



INDIAN AGRICULTURAL
RESEARCH INSTITUTE, NEW DELHI.

56377

३५

I. A. R. I. 6.

MOIPC SI -6 AR/51 -7-7-51 -10,000.

ANNALS OF THE NEW YORK ACADEMY OF SCIENCES
Volume 51, Art. 1. Pages 1-11

Editor
ROY WALDO MINER

Isidor I. Rabi, President
LOTHAR SÄTHL

NEWER SYNTHETIC ANALGESICS

BY

M. L. TANTER, H. J. ANSLINGER, R. C. BATTERMAN,
R. N. BIFTER, S. BOGOLUB, K. K. CHEN, N. B. EDDY,
L. GLASSMAN, B. B. HERSHENSON, S. A. HURSH,
R. W. HOUDÉ, H. ISBELL, E. H. JENNEY,
J. S. LADUE, L. C. MILLER, E. I. NELSON,
C. C. PEPPIER, L. H. RASMUSSEN, I. A. ROVINSKY,
M. H. SEEVERS, L. F. SMALL, R. R. SONNENSCHIEIN,
AND F. P. YONKMAN



NEW YORK
PUBLISHED BY THE ACADEMY
November 1, 1948

(LYCEUM OF NATURAL HISTORY, 1817-1876)

OFFICERS, 1948

President

HARDEN F TAYLOR

Vice-Presidents

GEORGE B PEGRAM

HANS MOLITOR

Recording Secretary

Corresponding Secretary

GEORGE B. PEGRAM

ATHELSTAN F. SPILHAUS

Treasurer

E. CLIFFORD WILLIAMS

Editor

ROY WALDO MINER

Elected Councillors

1946-1948

HANS T CLARKE CHARLES GLEN KING

1947-1949

WILLIAM BROWN BELL HARRY A. CHARIPPER CARYL P. HASKINS

1948-1950

ROSS F. NIGRELLI VICTOR K. LAMER FREDRICK F. YONKMAN

Finance Committee

ADDISON WEBB, Chairman CHARLES N. FREY JOHN A. ZELLERS

Executive Director

BUNICE THOMAS MINER

OFFICERS OF SECTIONS

GEOLOGY AND MINERALOGY

HENRY D. THOMPSON, Chairman GEORGE F. ADAMS, Secretary

STOCKS

CHARLES R. SCHROEDER, *Chairman* **ELI D. GOLDSMITH**, *Secretary*

PSYCHOLOGY

CLAIRETTE F. ARMSTRONG, *Chairman*; LAWRENCE E. AUST, *Secretary*

1999

WILLIAM DUNCAN STRONG, Chairman CHARLES W. WAGLEY, Secretary

PHYSICS AND COSMOLOGY

SEYMOUR A. KORFF, Chairman LEO MEDICOVSKY, Secretary

POEATOGRAPHY AND METEOROLOGY

BERNARD HAUROWITZ, Chairman JOHN C. BARTON, Secretary

[illegible]

ANNALS OF THE NEW YORK ACADEMY OF SCIENCES

Volume 51, Art. 1. Pages 1-174

November 1, 1948

NEWER SYNTHETIC ANALGESICS*

Consulting Editor: M. L. TAINTER

CONTENTS

	PAGE
Pain. By M. L. TAINTER	3
Chemistry of Natural and Synthetic Analgesics. By LYNDON F. SMALL	12
Experimental Methods for Studying Analgesia. By CARL C. PFEIFFER, RALPH R. SONNENSCHIEIN, LIONEL GLASSMAN, ELIZABETH H. JENNEY, AND SALLY BOGOLUB	21
A Critique of Analgesic Testing Methods. By LLOYD C. MILLER	34
Pharmacology of Metopon and Other New Analgesic Opium Derivatives. By NATHAN B. EDDY	51
Pharmacology of Demerol and its Analogues. By FREDRICK F. YONKMAN	59
Pharmacology of Methadone and Related Compounds. By K. K. CHEN	83
Animal Experimentation in Studying Addiction to the Newer Synthetic Analgesics. By M. H. SEEVERS	98
Methods and Results of Studying Experimental Human Addiction to the Newer Synthetic Analgesics. By HARRIS ISBELL	108
The Importance of Addiction to the Newer Synthetic Analgesics in Human Therapy. By ROBERT C. BATTERMAN	123
The Federal Food, Drug and Cosmetic Act and New Analgesic Drugs. By ERWIN E. NELSON	130
Regulatory Problems of the New Analgesics under the Narcotic Law. By H. J. ANSLINGER	134
Methadone in Internal Medicine. By RAYMOND N. BIETER AND STANTON A. HIRSH	137
Use of New Synthetic Analgesics in Surgery. By E. A. ROVENSTINE	145
Use of the New Synthetic Analgesics in Obstetrics. By BERT B. HERSHENSON	151
Preliminary Experiences in the Use of Some of the Newer Analgesics in the Relief of Pain Due to Cancer. By R. W. HOEDE, L. H. RASMUSSEN, AND J. S. LADUE	161

* This series of papers is the result of a Conference on Newer Synthetic Analgesics held by the Section of Biology of The New York Academy of Sciences on May 14 and 15, 1948.

Publication made possible by contributions from Abbott Laboratories; Burroughs-Wellcome, Inc.; Ciba Pharmaceutical Products, Inc.; Eli Lilly Laboratories; E. R. Squibb and Sons; G. D. Searle Company; Hoffmann-LaRoche, Inc.; J. T. Baker Chemical Company; Lederle Laboratories Division, American Cyanamid Company; Merck and Company, Inc.; Sharp and Dohme, Inc.; Smith, Kline, and French Laboratories; Sterling-Winthrop Research Institute; The Upjohn Company; and the General Funds of The New York Academy of Sciences.

Copyright, 1948, by The New York Academy of Sciences

THE MERRYMOUNT PRESS, BOSTON, U. S. A.

PAIN

By M. L. TAINTER

Sterling-Winthrop Research Institute, Rensselaer, New York

... PAIN IS PERFECT MISERY, THE WORST
OF EVILS; AND, EXCESSIVE, OVERTURNS
ALL PATIENCE.

SO cried Milton in *Paradise Lost*¹; but the somber concern of the poet three hundred years ago, unfortunately, is still too much with us even today. Hence, the urge that draws us together in this conference to scrutinize the achievements of the recent past and to cast exploratory glances into the near future when, we hope, pain will be so completely controllable as to be no longer a major cause of anguish or suffering.

The history of man's continuous struggle to alleviate pain extends back through the centuries to earliest man. The story begins as thin threads of legend, fact, and fancy winding their way through the early ages in mythological writings, through the cultures and religious rituals of the ancient Hebrews, Egyptians, Greeks, and Romans, perilously surviving the Dark Ages, to emerge into the great modern era of synthetic analgesics, anesthetics, and narcotics. These developments are truly monuments to the creative genius of man.

The long story of the origin of modern analgesic drugs may be divided into four major eras. The first of these was characterized by man's discovery and use of natural plant products; the second, by the isolation of the pure alkaloids from these natural sources and their first use as therapeutic agents. The third great era of progress consisted of the development of organic chemistry with the consequent introduction of the first synthetic analgesic drugs. The fourth, or modern period, has been characterized by the development of quantitative pharmacological methods of testing making possible the systematic development of synthetic drugs which are even more potent than those produced by nature.

The first of these great eras began with primitive man's groping in the darkness of superstition and early religious mysticism. Probably the earliest attitude of our aboriginal forefathers toward pain was to consider it as a demon or evil spirit. Efforts were made to appease, or frighten away, these pain demons with rings worn in the ears and nose, or with amulets, tiger claws, talismans, and similar charms to ward off the evil spirits of pain. In addition, the bodies were often tattooed or scarified to frighten away the demons. Gradually, these procedures were replaced by the medicine man, conjurer or shaman of the tribe. He muttered magic incantations and wrestled and fought with the invisible pain demons, and then later in history gradually began to administer natural remedies selected through empirical observations. As the centuries flowed on, the medicine man was slowly replaced by the priest, who was

considered the servant of the gods. Along with the natural remedies, the priest relied on prayer, and the fate of the patient was thus entrusted to a higher power. The classic example of this era was the blond goddess Agamede of the Greeks who was considered to wield exclusive power over the demons of illness and pain.

Numerous examples of the intertwining of religious lore with medicine might be cited from the Bible. The story of the creation of Eve in the second chapter of Genesis² may be considered as an allegorical description of an early painless operation:

And the Lord God caused a deep sleep to fall upon Adam, and he slept; and He took one of his ribs, and closed up the flesh instead thereof.

With the birth of Christianity, a new concept of the relief of pain developed, based on divine healing through laying-on of hands and prayer. One of the primary tasks of the Son of God and his followers was the banishment of the suffering of all mankind. Therefore, the church of the Early and Middle Ages devoted much attention to the healing ministrations of its clergy.

The origin of the medicinal use of herbs is lost in the obscurities of antiquity. It is probable that early man, experimenting with various plants as foods, discovered that some of them were efficacious in assuaging pain and curing disease. It has been suggested that the early herbalist, finding that his skill brought him special prestige and profit, surrounded his knowledge with mystery and cant, thus identifying himself with the gods and increasing his importance and success. At any rate, the incantations, the amulets, and the rituals performed to call forth the merciful intervention of the gods were indispensable adjuncts to the prescriptions concocted by early physicians. The mystic herb concoctions and other professional secrets were handed down by the sorcerers, magicians, and sacrificial priests so that from this arose much of the natural science among the peoples of ancient Babylon, Egypt, Israel, Eastern Asia, Greece, and Rome.

The "divine drugs of sleep," and syrups that produced insensibility, were prominent in all ancient cultures. The earliest records recount legends on the effects of such plants as the poppy, mandragora, hemp, and henbane, whose narcotic properties gave comfort to all sufferers. The writings of Hippocrates, Pliny, Dioscorides, and others contain many references to the production of insensibility to pain by the use of alcohol, alcoholic extracts of opium, or the administration of mandragora or other stupefying drugs.

The first written records mentioning remedies for pain are probably those contained in ancient Babylonian clay tablets said to date back to 2250 B.C. One of these describes a remedy for the pain of dental caries which, according to Prinz,³ was a cement consisting of henbane seeds in powdered form mixed with gum mastic, which was applied to the cavity in the tooth. The Ebers Papyrus, reputedly written in 1552 B.C., con-

tains a collection of early Egyptian prescriptions and formulae, some of which contained opium, henbane, and mandragora. A prescription consisting of opium, coriander, wormwood, juniper, and honey is described, which, according to legend, was a remedy of divine origin which Isis prescribed for Ra's headaches.

The Iliad and the Odyssey of the Greek poet Homer, presumably written during the ninth century B.C., both contain references to the use of analgesic drugs for relief from pain. In the Iliad, Erypylus, wounded in battle, makes this request of Patroclus:

With lukewarm water wash the gore away,
With healing balms the raging smart allay,
Such as sage Chiron, sire of Pharmacy,
Once taught Achilles, and Achilles thee.

Patroclus carried out his request and:

Cut out the biting shaft; and from the wound
With tepid water cleansed the clotted blood;
Then pounded in his hands, the root applied
Astringent, anodyne, which all his pain
Allay'd; the wound was dried, and stanch'd the blood.

In the Odyssey, Homer records that Helen, the daughter of Zeus, prepared a drug, possibly opium, dissolved in wine, to sleep off grief and anger and to forget pain:

Presently she cast into the wine whereof they drank, a drug to lull all pain and anger, and bring forgetfulness to every sorrow.

According to Macht,⁴ the first authentic reference to the milky juice of the poppy is in the writings of Theophrastus in the third century B.C. The mandragora plant, or mandrake, is also described by Theophrastus in his *Enquiry into Plants*,⁵ as follows:

The leaf of mandrake, they say, is useful for wounds, and the root for erysipilas, when scraped and steeped in vinegar, and also for gout, for sleeplessness, and for love potions.

Scribonius Largus in his *Compositiones Medicamentorum*, written about 40 A.D.,⁶ described the method for procuring opium and pointed out that the opium of commerce was derived from the capsules and not from the foliage of the plant.

Probably one of the first references to analgesic pills was made by Celsus⁷ in his *De Medicina*, written during the first century A.D., which reads:

Pills are also numerous, and are made for various purposes. Those which relieve pain through sleep are called anodynes; unless there is overwhelming necessity, it is improper to use them; for they are composed of medicaments which are very active and alien to the stomach. There is one, however, which actually promotes digestion; it is composed of poppy-tears and galbanum, myrrh, castory, and pepper. . . . Another, worse for the stomach, but more soporific, consists of mandragora, celery seed, and hyoscyamus seed, which are rubbed up after soaking in wine.

Also in the first century A.D., Dioscorides, who served as a surgeon in

the armies of Nero, wrote so extensively on drugs that he has often been called the Father of Pharmacy. His *Greek Herbal*, which has been described as the first materia medica, contains rather detailed discussions of the botany and medicinal uses of the opium poppy, the mandragora plant, henbane, and hemp. As can be seen from the following lines taken from the section on *Mekon agrios* & *emeros*, as translated by John Goodyer in 1655,^{3a} Dioscorides plainly distinguished between the juice of the capsules of the opium poppy and extracts from the entire plant, and described exactly how the capsules should be incised.

But it behooves them that make Opium after that the dew is dried away to scarify about the asterisk with a knife, so that it do not pierce into ye inside, & from ye sides of the head make straight incisions in ye outside, & to wipe off a tear that comes out with ye finger into a spoone, & again to return not long after, for there is found another (tear) thickened, & also on the day after. But it must be beaten in a mortar, & be laid up when made into trochisks, but in ye cutting it, you must stand back, to ye end that the liquor be not wiped away of your clothes.

Ye seed of the black Poppy beaten small is given to drink with wine for ye flux of ye belly, & ye womanish flux. It is applied unto such as cannot sleep with water upon ye forehead and the temples, but the liquor itself is more cooling & thickening & drying. A little of it taken as much as a grain of Ervum, is a pain easer & a sleep causer, & a digester, helping coughes, & Coeliacall affections. But being drank too much, it hurts, making men lethargicall & it kills.

Dioscorides also gives a good description of the drug mandragora^{3b}:

And some do seeth the roots in wine to thirds, & straining it set it up. Using a Cyathus of it for such as cannot sleep, or are grievously pained, & upon whom being cut, or cauterized they wish to make a not-feeling pain. Ye juice being drank ye muchness of ye quantity of 2 Oboli with Melicrate, doth expel upward Phlegm, and black choler, as Ellebore doth, but being too much drank it drives out ye life.

We may infer from these and other statements of Dioscorides that the collection of opium was definitely a branch of industry in Asia Minor during the first century A.D.

At about the same time, Pliny the Elder described mandragora in language that leaves no question as to its use as an analgesic³:

Yet it may be used safely ynough for to procure sleepe, if there be a good regard had in the dose. . . . Also it is an ordinarie thing to drinke it against the poyson of serpents: Likewise before cutting, cauterizing, pricking or launcing of any member, to take away the sence and feeling of such extreame cures.

The dosage forms of opium, mandragora, henbane, and hemp used at this period were crude extracts, syrups, powders, pill masses, poultices, ointments, and various complex concoctions dissolved in wine and other alcoholic beverages.

Galen, in the second century A.D., spoke enthusiastically of the virtues of opium and mandragora. He mentions the use of mandragora to paralyze sensation and motion, the use of the tears of the poppy in a toothache prescription, and other "syrups of sleep" to relieve suffering.

The teachings of Galen dominated medicine for more than the follow-

ing thousand years, and during this time few advances were made in the field of analgesic drugs.

As the ignorance and superstition of the Dark Ages settled over Europe, the center of medical knowledge shifted to the Arabs. Avicenna, the dominant figure of the Arabian school of medicine, codified all available medical knowledge in his *Canon of Medicine*, and this, translated into Latin, became the authoritative medical textbook of Europe for six centuries. According to Robinson,¹⁰ Avicenna discussed fifteen types of pain: boring, compressing, corrosive, dull, fatigue, incisive, irritant, itching, pricking, relaxing, stabbing, tearing, tension, and throbbing. Both Rhazes and Avicenna advocated the use of opium, mandrake, and henbane, but these were still used in the ancient prescription forms of crude extracts, syrups, powders, pill masses, and other complex concoctions of the crude drugs.

The somniferant sponge became quite popular in Europe during the latter part of the Middle Ages as the medical knowledge of the Arabs gradually crept back into Southern Europe. The *Antidotarium* of Nicholas of Salerno¹¹ contains the following description of the preparation and use of the "*Spongia Somnifera*":

Take . . . of opium thebaicum, juice of hyocyamine, unripened berry of the blackberry, lettuce seed, juice of hemlock, poppy, mandragora, ivy. . . . Put these all together in a vessel and plunge therein a new seasponge just as it comes from the sea, taking care that fresh water does not touch it. And put this in the sun during the dog-days until all the liquid is consumed. And when there is need, dip it a little in water not too warm, and apply it to the nostrils of the patient, and he will quickly go to sleep. When, moreover, you want to awaken him, apply juice from the root of the fennel and he will soon bestir himself.

The origin of our modern tincture of opium or laudanum dates back to that famous physician of the Middle Ages, Philippus Aureolus Theophrastus Bombast von Hohenheim, commonly known as Paracelsus, who lived from 1490 to 1540, and who did much to overthrow the domination which Galen had exerted over medicine for centuries. Paracelsus probably applied the term "laudanum" (something to be praised) to several different preparations, all of which contained opium as the basic constituent. The laudanum of the early London pharmacopoeias contained opium, wine, and other ingredients. The principal liquid preparation of opium used in England in the next century was Sydenham's laudanum described by him about 1670. Near the same time, another preparation, called Rousseau's laudanum, was much in vogue on the continent.

The so-called "black-drop" was another celebrated opiate of the eighteenth century. This preparation was devised by one Edward Runstall of Auckland and was also known as Lancaster or Quaker's black-drop. A formula for its preparation is given by Macht⁴ as follows:

Opium, $\frac{1}{2}$ pound; verjuice, 4 pints; nutmegs, $1\frac{1}{2}$ ounces; saffron, $\frac{1}{2}$ ounce. Boil, add two spoonfuls of yeast and set in a warm place for six to eight weeks; then decant, filter and put in bottles.

This preparation was about three times as strong as laudanum and was the forebear of the modern English *Acetum Opii*.

Another familiar household remedy, "paregoric," originated with the *Elixir Asthmaticum* of Le Mort, who was a professor of chemistry at the University of Leyden from 1702 to 1718. The word paregoric is derived from the Greek for "soothing" or "consoling." The London Pharmacopoeia of 1746 describes this preparation under the name *Elixir Paregoricum*, and in 1888 the official name became *Tinctura Opii Camphorata*.

Our modern opium pills, or *Pilulae Opii*, are descended from the old English *Pilulae Saponis Co.*, or *Pilulae Saponacae*, which in turn are an adaptation of the nostrum known as Matthew's Pills or Starkey's Pills. The famous Dover's powder of Thomas Dover of Robinson Crusoe fame, consisting of powdered opium and ipecac, might also be mentioned in passing.

Thus, at the end of the eighteenth century, we find that opium, mandragora, and henbane were the principal drugs employed for the alleviation of pain. With the exception of a few new prescription forms, these natural drugs were still used much the same as the Egyptians, Greeks, and Romans had used them two thousand years before. The hundreds of years of the Middle Ages had contributed little to man's search for a more effective analgesic agent. Despite its manifold contributions to the other sciences, the Renaissance had made a disappointing record in its contributions to the relief of pain. The active principles (alkaloids) of the natural drugs were still unknown and it was therefore necessary to administer the natural drugs themselves in crude form. The potency and action of these crude extracts and concoctions were unpredictable since there were no methods for standardizing the dosage, and no way to control the strength of the raw products used in their formulation. Therefore, progress in the field of analgesia reached an early stalemate and had to wait to be revitalized by new ideas and concepts.

The new idea, which initiated the second great era in analgesia, was introduced by the young German apothecary, Friedrich Wilhelm Adam Sertürner. Although a relatively uneducated apothecary's assistant, Sertürner conceived the idea of isolating the active constituent of opium, developed sufficient skill to obtain the product he desired, and then with true scientific fervor proceeded to carry out, not only preliminary pharmacological testing on animals, but also clinical applications upon himself and his poor unsuspecting friends. His first paper published in 1806¹² was ignored, but eventually the importance of getting pure crystalline drugs out of previously crude and uncertain mixtures was recognized, and led to the new science of alkaloidal chemistry. Sertürner first called his new material *Principium Somniferum*, then in 1817¹³ named his new alkaloid "morphine" after Morpheus, the Greek god of dreams, who was one of the many sons of Hypnos, god of sleep. It is perhaps indicative of the importance of analgesia that the first alkaloid to be isolated was morphine.

Sertürner's work quickly led to the isolation of other opium alkaloids such as codeine in 1832 by Robiquet,¹⁴ thebaine by Pelletier in 1835,¹⁵ and papaverine by Merck in 1848.¹⁶ It ushered in a most prolific scientific period characterized by the extraction and purification of most of the important alkaloids from their natural vegetable sources. The isolation of quinine, strychnine, cocaine, nicotine, atropine, brucine, emetine, narceine, veratrine, caffeine, theophylline, etc., followed quickly on the heels of Sertürner's pioneering discoveries. These natural alkaloids quickly found wide applications in medicine which have largely persisted to the present day.

The foundation for an entirely new era, the third, in analgesic science was laid in 1828, when the young German chemist, Friedrich Wöhler, reported his synthesis of urea from inorganic materials.¹⁷ Urea had been known for years as a product of animal metabolism and as a constituent of urine. Its synthesis by purely chemical means broke down the old distinction between organic and inorganic chemistry, thus causing the overthrow of the theory that a "vital force" present in living material was necessary for the formation of organic compounds. His magnificent discovery paved the way for this third great era in the development of analgesics, namely, the period of the first successful analgesic drugs made by chemical synthesis.

This period began with the salicylates in 1827, when Leroux discovered salicin, derived from willow bark.¹⁸ Shortly after, salicylic acid was made from salicin by Piria in 1838,¹⁹ from oil of wintergreen by Cahours in 1844,²⁰ and synthesized from phenol in 1860 by Kolbe and Lautemann.²¹ Sodium salicylate was used for the first time by Buss in 1875,²² and phenylsalicylate in 1886 by Nencki.²³ In 1899, Dreser²⁴ discovered acetylsalicylic acid which he called aspirin from the German word "*Spirsäure*," meaning salicylic acid. This is still the most popular analgesic in terms of volume of use. Acetanilid was introduced into medicine as an antipyretic by Cahn and Hepp in 1886,²⁵ under the name Antifebrine. Acetophenetidin, or Phenacetin, followed in 1887.²⁶ It was not long before their importance as antipyretics was overshadowed by their real effectiveness as analgesics against many kinds of pain.

In 1884, fifty-six years after Wöhler, Ludwig Knorr, while searching for a synthetic substitute for quinine, prepared Antipyrine, a compound whose analgesic powers probably have not been fully utilized.²⁷ Knorr's synthesis of Antipyrine marked the beginning of the famous German drug industry and ushered in Germany's forty-year dominance of the synthetic drug and chemical field. Aminopyrine (Pyramidon), a still more potent compound closely related to antipyrine, was synthesized a few years later in 1896 by Knorr and Stolz.²⁸ This remained the most powerful analgesic of synthetic origin until recent times.

In the fourth, or modern era of analgesia an increased understanding of organic chemistry has been coupled with new and quantitative methods of testing analgesic drugs. These advances have finally made it possible

to synthesize compounds which are better than those found in nature. The development of quantitative tests for analgesic power has been the crucial element in this era, since through these methods it became possible for the first time to follow modifications in potency which accompanied systematic changes in the chemical molecule.

The lead was taken in 1929 by Small, Eddy, and their colleagues under the sponsorship of the Committee on Drug Addiction of the National Research Council. Their objective was to find a potent analgesic derived chemically from the opium group, which would produce the beneficial effects of morphine without addiction or undesirable side actions. The most important drug to evolve from these studies, Metopon, will undoubtedly be given extended consideration in this Conference.

Finally, in 1939, during a systematic search for antispasmodic agents, Eisleb and Schaumann²⁹ discovered that phenylpiperidine esters had sufficiently high analgesic power to be practical therapeutic agents for the control of pain. The most generally known of these today is the compound meperidine or isonipecaine. Introduction of this drug was immediately followed by its widespread use, since the new compound controlled pain and smooth-muscle spasm in a highly efficacious manner without many of the disadvantages associated with morphine.

This recent demonstration that synthetic chemistry could produce compounds equal or superior in therapeutic value to the natural opiates has stimulated vigorous research, which has resulted in the synthesis of compounds of great scientific interest and probable practical value. Among these are several analogs of meperidine and a series of compounds with a different type of chemical structure, as represented by methadone and isomethadone. Several of these new drugs are even more potent in their analgesic effects than morphine.

At long last, the synthetic chemist and the pharmacologist, with his modern concepts of quantitative pharmacology, have teamed up to evolve new drugs which already at the beginning of an era surpass the best that nature has to offer, and which are notable improvements over the preparations which were the sole pain-relievers for tens of centuries.

Today, we are assembled to summarize and to discuss these recent developments, and to gain stimulus and insight into the future. If there is any such thing as "fecundity by association," it is not too much to expect that, in this Conference, there may be conceived a new major era in pain control, which ultimately will result in the birth of new modes of analgesia, effective beyond all the bounds of our present imagination.

Bibliography

1. MILTON, J. *Paradise Lost* vi: 462-464.
2. *Holy Bible*. Genesis iii: 21.
3. PRINZ, H. 1924. Local anesthesia as applied to dentistry. In: GWATHMEY, J. T.: *Anesthesia*. Ed. 2: 535. Macmillan Co. New York.
4. MACHET, D. I. 1915. The history of opium and some of its preparations and alkaloids. *J. A. M. A.* 64: 477-481.

5. THEOPHRASTUS. *Enquiry into Plants*. (Quoted from ROBINSON.¹⁰)
6. SCRIBONIUS LARGUS. 1786. *Compositiones Medicamentorum*. (Quoted from MACHT.⁴)
7. CELSUS. *De Medicina*. (Quoted from ROBINSON.¹⁰)
- 8a. *The Greek Herbal of Dioscorides*: 456. (Translated by John Goodyer in 1655.) 1934. Oxford University Press.
- 8b. *Ibid.*: 473.
9. PLINY: Cited by KEYS, T. E. 1945. *The History of Surgical Anesthesia*. Schuman. New York.
10. ROBINSON, V. 1946. *Victory over Pain*. Schuman. New York.
11. NICOLAUS SALERNITANUS. 1471. *Antidotarium*, f.33 b. Nicolaus Jensen, Venice. (Quoted from KEYS.⁹)
12. SERTÜRNER. 1806. Darstellung der reinen Mohnsäure (Opiumsäure) nebst einer chemischen Untersuchung des Opiums mit vorzüglicher Hinsicht auf einen darin neu entdeckten Stoff und die dahin gehörigen Bemerkungen. J. Pharm. Aerzte, Apoth. Chem. 14: 47. (Quoted from KRUEGER, H., N. B. EDDY, & M. SUMWALT. 1943. *The Pharmacology of the Opium Alkaloids*. Suppl. 165, Pub. Health Rep.)
13. SERTÜRNER. 1817. De la morphine et de l'acide méconique, considérés comme parties essentielles de l'opium. Ann. chim. phys. [2.s.] 5: 21. Über das Morphinum, eine neue salzfähige Grundlage und die Mekonsäure, als Hauptbestandtheile des Opiums. Ann. Phys. (Gilbert) 55: 56. Über eins der fürchterlichsten Gifte der Pflanzenwelt, als ein Nachtrag zu seiner Abhandlung über die Mekonsäure und das Morphinum, mit Bemerkungen, den aciden Extractivstoff des Opiums und seine Verbindungen betreffend. Ann. Phys. (Gilbert) 57: 183. (Quoted from KRUEGER, H., N. B. EDDY, & M. SUMWALT—see ref. 12.)
14. ROBQUET. 1833. Neue Beobachtungen über die vorzüglichsten Produkte aus dem Opium. Ann. 5: 82-111.
15. PELLETIER, J. 1835. Neue Untersuchungen, die Geschichte und die näheren Bestandtheile des Opiums betreffend. Ann. 16: 27-54.
16. MERCK, G. 1848. Vorläufige Notiz über eine neue organische Base im Opium. Ann. 66: 125-128.
17. WÖHLER, F. 1828. Ueber künstliche Bildung des Harnstoffs. Ann. Phys. Chem. 12: 253-256.
18. LEROUX: Cited by PETIT, L. H. 1876. De l'emploi de la salicine, de l'acide salicylique et du salicylate de soude dans la thérapeutique médicale. Bull. gén. thérap. 91: 454-465.
19. PIRIA, R. 1838. Recherches sur la salicine et les produits qui en dérivent. Ann. chim. phys. 69: 281-325.
20. CAHOURS, A. 1844. Recherches chimiques sur le salicylate de méthylène et l'ether salicylique. Ann. chim. phys. [3.s.] 10: 327-369.
21. KOLBE, H., & E. LAUTEMANN. 1860. Ueber die Constitution und Basicität der Salicylsäure. Ann. Chem. Pharm. 115: 157-206.
22. BUSS, C. E. 1875. Ueber die Anwendung der Salicylsäure als Antipyreticum. Deut. Arch. klin. Med. 15: 457-501.
23. NENCKI, M., & F. NACHFOLGER. 1886. Verfahren zur Darstellung der Salicylsäureester der Phenole und Naphtole, genannt "Salole." Fdrl.: 237.
24. DRESER, H. 1899. Pharmakologisches über Aspirin (Acetylsalicylsäure). Arch. ges. Physiol. 76: 306.
25. CAHN, A., & P. HEPP. 1886. Das Antifebrin, ein neues Fiebermittel. Centralbl. klin. Med. 7: 561-564.
26. HINSBERG, O., & A. KAST. 1887. Ueber die Wirkung des Acetphenetidins. Centralbl. med. Wissensch. 25: 145-148.
27. KNORR, L. 1884. Ueber die Constitution der Chinizenderivate. Ber. 17: 2032-2049.
28. KNORR, L., & F. STOLZ. 1896. Ueber das 4-Amidoantipyrin. Ann. 293: 58-69.
29. EISLEB, O., & O. SCHAUMANN. 1939. Dolantin, ein neuartiges Spasmolytikum und Analgetikum (Chemisches und Pharmakologisches). Deut. med. Wochenschr. 65: 967.

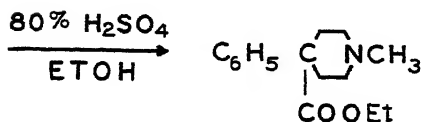
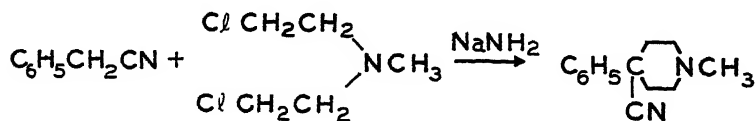
CHEMISTRY OF NATURAL AND SYNTHETIC ANALGESICS

By LYNDON F. SMALL

National Institute of Health, Bethesda, Maryland

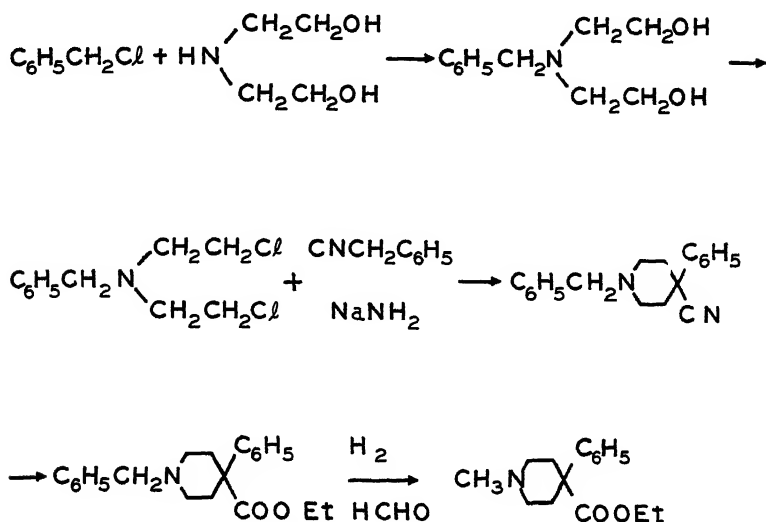
THERE has been a marked tendency, in recent years, to explain the analgesic action of some of the newly developed synthetic drugs in terms of certain structural features present in the morphine molecule: in particular, a specific positional relationship of a tertiary nitrogen atom to a quaternary carbon and an aromatic nucleus. It is notable that most of these hypotheses have been evolved *after* the analgesic action of the drugs was discovered, and that they tacitly ignore many instances where the key structure is present but little if any analgesic action is observed. The value of such speculation will be first apparent when new types of analgesics are discovered as a result of its application. As far as can be ascertained, both of the most prominent drugs under discussion were discovered in a search for spasmolytic agents.

The first of these, Demerol, Dolantin, pethidine, meperidine, or isonipecaine was developed by Eisleb¹ in a study of condensation reactions of benzyl cyanide in the presence of sodamide. In the original (journal) description, benzyl cyanide and *bis*-(2-chloroethyl) methylamine were condensed, and the resulting nitrile subjected to hydrolysis and esterification in a single step.



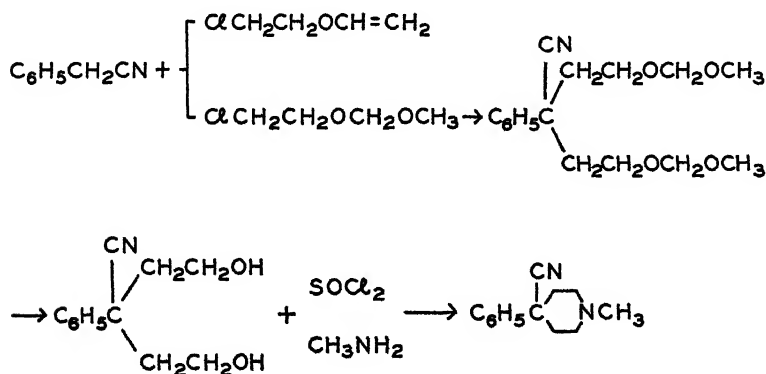
Because of the disagreeable properties of the *bis*-(chloroethyl)amine, a powerful vesicant, efforts have been made to circumvent its use. According to the O. P. B. report,² the I. G. Farbenindustrie process benzylated diethanolamine, the product was then chlorinated, and condensed with benzyl cyanide as in the Eisleb method. The resulting nitrile was converted to the ethyl ester, which was then catalytically deben-

zylated, and N-methylated with formaldehyde in a continuation of the reduction process.

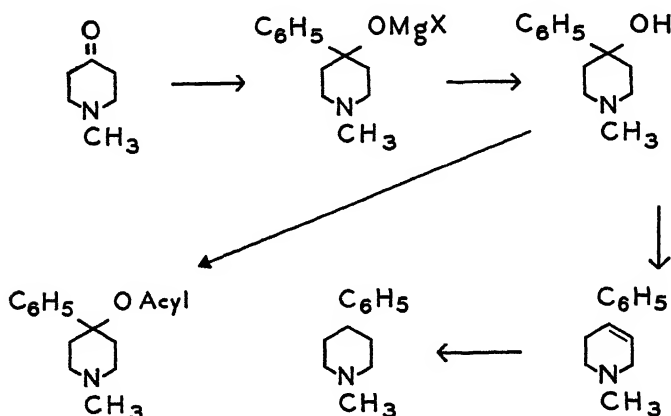


In the preparation of Bemidon, the *m*-hydroxyphenyl analog of Demerol, *m*-methoxybenzyl cyanide is used in the sodamide condensation, and simultaneous splitting of the methoxyl and hydrolysis of the nitrile group are accomplished with 66 per cent hydrobromic acid.²

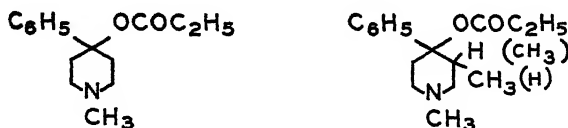
A variation of the Demerol synthesis developed by Bergel *et al.*³ utilizes the condensation of two moles of the methoxymethyl or vinyl ethers of 2-chloroethanol with benzyl cyanide. The labile ether linkage is then split by mild hydrolytic agents, and piperidine ring-closure performed by chlorination and reaction with methylamine.



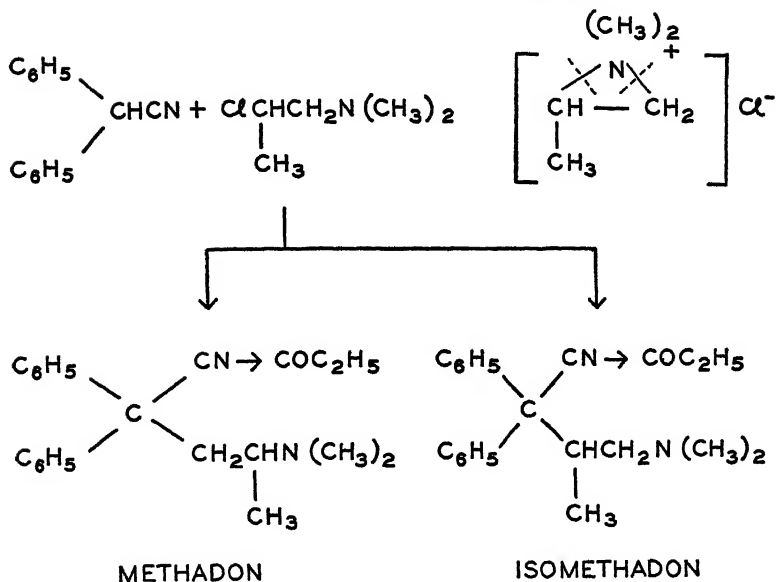
Although Demerol and Bemidon seem to be the most active of the piperidinecarboxylic ester types, they do not exhaust the potentialities of piperidine analgesics. A number of comparable ketones were investigated by the I.G.³ and found to be without particular promise. A different type of piperidine ester, studied independently by Jensen *et al.*⁴ and the Hoffmann-La Roche group,⁵ appears to be more active than the Demerol series. These are derived from N-methyl-4-phenyl-4-hydroxypiperidine. They are prepared by the action of Grignard reagents or organolithium compounds on 4-piperidones, followed by acylation of the organometallic complex. This is preferable to isolation of the free hydroxy compounds, which show considerable tendency to go over into unsaturated systems on acylation.



The most active of these appear to be the 4-propionyloxy and 3-methyl-4-propionyloxy derivatives, which are reputed to surpass morphine in analgesic power.⁶

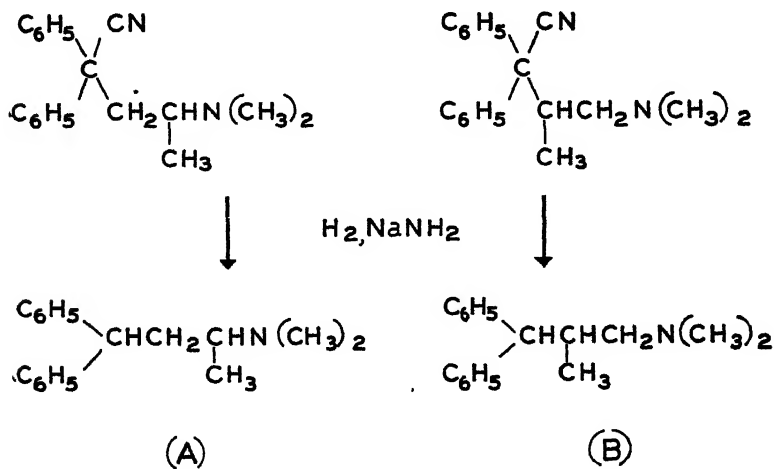


Since publication of the O. P. B. report⁷ on amidon (10820, methadone, Dolophine, Adanon, AN-148, Physeptone), there has been widespread interest in its chemistry and mechanism of formation. The condensation of diphenylacetonitrile and 2-chlorodimethylaminopropane in the presence of sodamide, or potassium *t*-butoxide, leads to a mixture of isomeric nitrile bases, the chloroamine apparently reacting through a cyclic quaternary ammonium form.^{8,9}



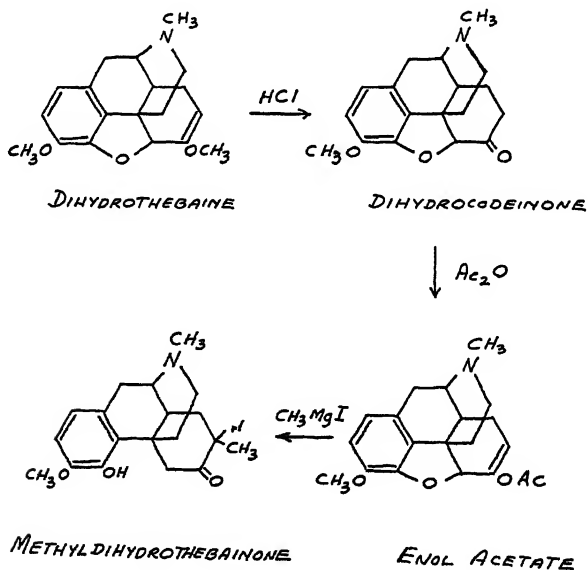
Treatment of this nitrile mixture with ethylmagnesium bromide gives methadone, and a ketimine which can be hydrolyzed with difficulty to isomethadone.¹⁰

The structures of these isomers have been demonstrated in two different ways, utilizing the corresponding nitriles. Bockmühl⁸ replaced the CN group with hydrogen by boiling in benzene with sodamide, and identified the resulting products A and B by synthesis.



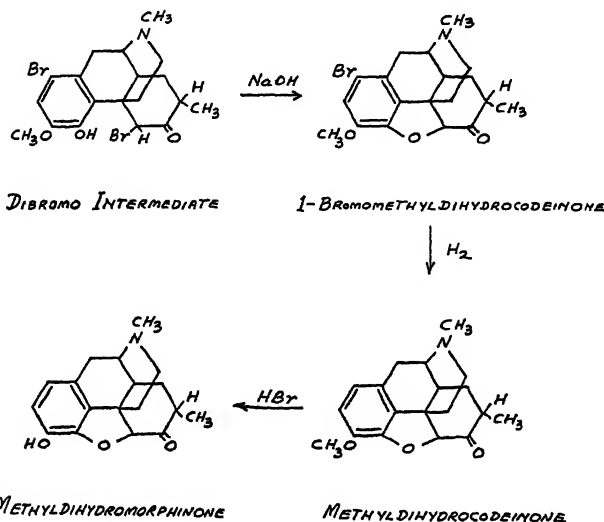
Methadone is not attacked by Clemmensen reduction, and in the modified Wolff-Kishner reaction of Huang-Minlon¹⁶ it undergoes an unusual type of split, in which the entire ketone chain is lost. The product, 3-dimethylamino-1,1-diphenylbutane, is probably formed through decarboxylation of an intermediate carboxylic acid. It was identified by exhaustive methylation and conversion to 1,1-diphenylbutane.¹⁵

Of the drugs of the morphine series, Metopon (methyldihydromorphinone) will first be considered. This compound was evolved by Small *et al.*¹⁷ in the course of a fundamental theoretical study of the remarkable reaction of thebaine with organomagnesium halides. The starting point of the synthesis may be either thebaine or codeine. Thebaine is hydrogenated to its dihydro derivative, which may be used as such, or hydrolyzed to dihydrocodeinone. The latter is also obtainable by catalytic rearrangement of codeine. Dihydrocodeinone or its enol acetate¹⁸ is brought into reaction with methylmagnesium iodide, which introduces a nuclear methyl group, probably at the 7-position, and opens the oxide ring.

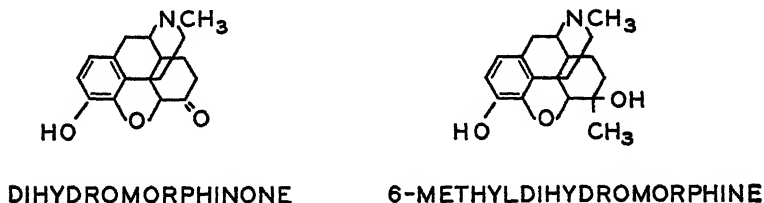


It is curious that the ketone group is untouched by excess reagent, but the morphine ketones are notorious for their indifference to $RMgX$ compounds, and the carbonyl may be in the form of an enol organometallic complex. For closure of the oxide ring, methyldihydrothebainone is treated with two moles of bromine, followed by alkali. The undesired bromine in the aromatic ring is removed by hydrogenation in the presence

of a hydrogen bromide acceptor (or the alkaloid itself, functioning as such), and finally the new ketone, methyl dihydrocodeinone, is demethylated at the methoxyl with one of the usual agents, to yield Metopon.



The abnormal behavior of the morphine ketones mentioned above has, until recently, made inaccessible the corresponding tertiary alcohols resulting from the Grignard reaction. It is found, however, that this sluggishness does not extend to reactions involving organolithium reagents.¹⁹ Dihydrocodeinone reacts instantly with methyllithium to give 6-methyldihydrocodeine, and, by the use of two moles of the reagent with dihydromorphinone, the corresponding 6-methyldihydromorphine is obtained. The latter is of interest for its prolonged analgesic action.



Application of this reaction to other morphine ketones, especially to methyldihydromorphinone (Metopon), may contribute to knowledge of the relationship of nuclear substituents to pharmacological action.

Bibliography

1. EISLER, O. 1941. New syntheses with sodamide. *Ber.* **74**: 1433-1430.
2. Office of the Publication Board, Department of Commerce, Washington, D. C. 1945. Pharmaceutical activities at the I. G. Farbenindustrie plant, Hoechst am Main. Report PB 981: 94-96.
3. BERGEL, F., A. L. MORRISON, & H. RENDERSNECHT. 1944. Synthetic analgesics. II. A new synthesis of Pethidine and similar compounds. *J. Chem. Soc.*: 265-267.
4. JENSEN, K. A., F. LUNDQUIST, E. REKLING, & C. G. WOLFFBRANDT. 1943. A new type of piperidine derivative with analgesic and spasmolytic activity. *Dansk. Tids. Farm.* **17**: 173-182.
1945. *Chem. Abstr.* **39**: 2506.
5. ZIERING, A., L. BERGER, S. D. HEINEMAN, & J. LEE. 1947. Piperidine derivatives. III. 4-Arylpiperidines. *J. Org. Chem.* **12**: 894-903.
6. ZIERING, A., & J. LEE. 1947. Piperidine derivatives. V. 1,3-Dialkyl-4-aryl-4-acyloxy-piperidines. *J. Org. Chem.* **12**: 911-914.
7. *Ref.* **2**: 96, 96-A, 97.
8. BOCKMUHL, M. 1946. Private communication to E. C. Kleiderer.
9. SCHULTZ, E. M., C. M. ROBB, & J. M. SPRAGUE. 1947. The reaction of aminoalkyl halides with diphenylacetone. The structure of amidone. *J. Am. Chem. Soc.* **69**: 188-189.
10. EASTON, N. R., J. H. GARDNER, & J. R. STEVENS. 1947. Some isomers of amidone and related compounds. *J. Am. Chem. Soc.* **69**: 976-977.
11. EASTON, N. R., J. H. GARDNER, M. L. EVANICK, & J. R. STEVENS. 1948. Some isomers of amidone and related compounds. *J. Am. Chem. Soc.* **70**: 76-78.
12. EASTON, N. R., J. H. GARDNER, & J. R. STEVENS. 1947. A new synthesis and confirmation of the structure of amidone. *J. Am. Chem. Soc.* **69**: 2941-2942.
13. THORP, R. H., E. WALTON, & P. OFNER. 1947. Optical isomers of amidone, with a note on isoamidone. *Nature* **160**: 605-606.
14. BRODE, W. R., & M. W. HILL. 1948. The optical resolution of amidone. *J. Org. Chem.* **13**: 191-193.
15. MAY, E. L., & E. MOSETTIG. 1948. Some reactions of amidone. *J. Org. Chem.* **13**: 459-464.
16. HUANG-MINLON. 1946. A simple modification of the Wolff-Kishner reduction. *J. Am. Chem. Soc.* **68**: 2487-2488.
17. SMALL, L., H. M. FITCH, & W. E. SMITH. 1936. The addition of organomagnesium halides to pseudocodeine types. II. Preparation of nuclear alkylated morphine derivatives. *J. Am. Chem. Soc.* **58**: 1457-1463.
18. SMALL, L., S. G. TURNBULL, & H. M. FITCH. 1938. The addition of organomagnesium halides to pseudocodeine types. IV. Nuclear substituted morphine derivatives. *J. Org. Chem.* **3**: 204-232.
19. SMALL, L., & H. RAPOPORT. 1947. Nuclear substituted morphine derivatives. *J. Org. Chem.* **12**: 284-292.

EXPERIMENTAL METHODS FOR STUDYING ANALGESIA

By CARL C. PFEIFFER, RALPH R. SONNENSCHN, LIONEL
GLASSMAN, ELIZABETH H. JENNEY, AND SALLY BOGOLUB

*Departments of Pharmacology and Clinical Sciences, University of Illinois College of
Medicine, and the Department of Anesthesia of St. Luke's Hospital, Chicago, Illinois*

SINCE pain is a subjective phenomenon, the testing of analgesic drugs in animals must, perforce, be limited to the determination of, and change in, their reactive thresholds. While this may correspond closely to their pain threshold, we have no way of determining for any species, except man, the difference between the reactive and the pain threshold. Because many drugs other than analgesics will also modify the reactive threshold, analgesic testing of new chemicals in animals must go hand in hand with a more extensive screening program to rule out such non-specific effects as hypothermia, methemoglobinemia, general shock, and curare-like action.

While TABLE 1 summarizes the many methods (see Goetzl¹ for an ex-

TABLE 1

METHODS FOR MEASUREMENT OF THE REACTIVE THRESHOLD IN ANIMALS

<i>A. Thermal</i>	<i>D. Chemical</i>
(1) Radiant heat	(1) Acids, potassium
<i>a.</i> Variable time	(2) Croton oil wheel
<i>b.</i> Variable heat	
(2) Contact heat—Reaction time	<i>E. Pharmacological</i>
<i>B. Electrical</i>	(1) Antipyretic effect
(1) Cutaneous	(2) Potentiation of depression or co-
(2) Tooth pulp	deine analgesia
	(3) Straub tail effect
<i>C. Mechanical</i>	(4) Action currents of nerves
(1) Toothed forceps	(5) Uricosuric effect
(2) von Frey hairs	
(3) Pressure	

tensive review) which have been utilized in the past to study analgesic and antipyretic drugs, only three methods with their modifications are now in general laboratory use. These are graded mechanical stimuli, graded electrical stimulation, and graded or timed heat stimuli.

Mechanical Methods. Von Frey hairs, made of horsehairs of various diameters and calibrated on a balance pan, have been used successfully by Seevers and Pfeiffer² to determine the degree and duration of analgesia produced by various opiates in man. Sensitive areas of the face must be used and multiple stimulation is required to assure several contacts of the needle with pain points. Under these conditions, the horsehairs are rapidly broken and must be replaced. To obviate this disadvantage,

SeEVERS³ has designed a mechanical algometer which allows a small weight to be moved along a beam so that a single needle may be weighted in a range of 0.10 to 20 grams. This simple apparatus deserves greater clinical trial for the study of changes in superficial pain (supain) in clinical patients.

The toothed forceps and also deep pressure have been used extensively in laboratory animals to elicit a squeak or a squeal response. While potent analgesics will undoubtedly reduce the squeak response, no one has, as yet, published a graded mechanical stimulus which, when used in animals, can be validated on a careful statistical basis.

Electrical Stimulation Methods. Various stimulators have been devised and used for the production of painful or irritative currents of high voltage and low amperage. When these stimuli are applied to the skin, the threshold varies widely from animal to animal and from one human subject to another. From the data of Macht and Macht,⁴ who studied calibrated inductorium stimuli applied to the scrotal region of rats, the ratio of the standard deviation to the mean is 58 per cent. In close comparison with these data from the rat are those of Lanier⁵ who applied condensor discharge stimuli to the skin of sixteen normal human subjects and found a wide variation (—80 to 300 per cent) in their normal pain thresholds. The ratio of the standard deviation to the mean of Lanier's data is 56 per cent, a remarkable coincidence since the Machts were studying the reactive threshold while Lanier was studying the pain threshold. These variations are much greater than thresholds determined by the radiant heat pain stimulus,⁶ and thus electrical stimulation of the skin can be discarded as a method of testing for analgesia unless better methods are found for reducing or standardizing skin resistance.

Electrical stimulation of the tooth pulp was first used in dogs by Koll and Reffert,⁷ and has been used by Ivy, Burrill and Goetzl⁸ in man where an induced current of calibrated voltage is applied by means of a pointed electrode to an occlusal metal filling while a large indifferent electrode is clasped by one hand under water or saline. The tooth pulp, presumably, contains only pain sensation so that only pain is elicited. It is possible to test both the lowest pain threshold (T_1) and the reactive threshold (T_2), which is the stimulus at which unbearable or severe pain occurs. In our limited experience, T_2 is more influenced by analgesic drugs while depressant drugs may have a greater effect on T_1 . Statistical evaluation of this method has been made by Harris,⁹ and also by Sonnenschein.¹⁰ The former found a low degree of variance if a single tooth was used in the same subject on a single day. Since analgesic testing is usually done in this manner, the method is very reliable. Care must be taken to apply the electrode to the same spot in the dry filling, and the subject must apply a constant finger area to the indifferent electrode. Sonnenschein studied nitrous oxide analgesia by the tooth-pulp

method and found that with fifteen subjects a mean rise of 25 per cent in T_1 pain threshold had a p value of less than 0.01, while a mean rise of 25 per cent in T_2 pain threshold in six subjects had a p value of less than 0.05. Since potent analgesics will produce a mean rise of more than 70 per cent in T_2 and 50 per cent in T_1 , the peak changes are undoubtedly statistically significant. We are, at present, comparing this method with the graded radiant heat method as applied to the blackened finger pad and nail bed. We find the tooth pain thresholds to be reliable and to be the most easily affected of any pain threshold. The tooth pain is aching in nature and is, thus, probably subserved by protopathic nerve fibers which we have called "deepain."

Radiant Heat-Pain Methods. These may be divided into three essential types which are: (1) constant contact heat for a variable period of time as applied to the feet of mice by the method of Woolfe and MacDonald¹¹; (2) constant radiant heat applied for a variable period of time until reaction occurs, which is the method of D'Amour and Smith¹² as applied to the rat's tail; and (3) variable radiant heat usually applied for 3 to 4 seconds to the skin of animals (Winder, Pfeiffer, and Maison¹³) or man (Hardy, Wolff and Goodell¹⁴), and the heat increased until pain or reaction occurs to the stimulus.

We have not used the contact heat method, but we understand that the end-point (the time when the mice start dancing or blowing on their paws) is difficult to time. It would seem that a modification of this method by the use of a sloping hot plate (similar to the sloping screen used in the insulin assay) might be advantageous in that the mice would start rolling down the hot plate when their feet became painful.

The method of constant radiant heat and variable time is successfully used in many commercial laboratories for the screening of new chemicals for possible analgesic action. When the rat's tail is used and the reaction time noted, great care must be taken to have the animals housed and tested in a constant temperature room because the temperature of the tail will vary with its environment, and also with the strain and health of the rat.

Winder, Pfeiffer, and Maison¹³ have applied variable radiant heat to the back of black guinea pigs, which species has a well-developed skin muscle (cutaneous maximus). In our experience, this method is applicable to the guinea pig, the dog, and the immature rabbit. Other species tested and found lacking in adequate skin muscles are the hamster, the mouse, and the rat. Andrews and Workman¹⁵ used the flank of the trained dog, which method has been found suitable in several commercial laboratories. When a stimulating time of 3 to 5 seconds is used and the threshold is approached from a high level, the standard error of the determination is exceedingly low (FIGURE 1). Ideally, five seconds is optimal, but the 3-to-4-second stimulus produces less eye fatigue in the operators. Several independent observers have been able to assay the analgesic effect of

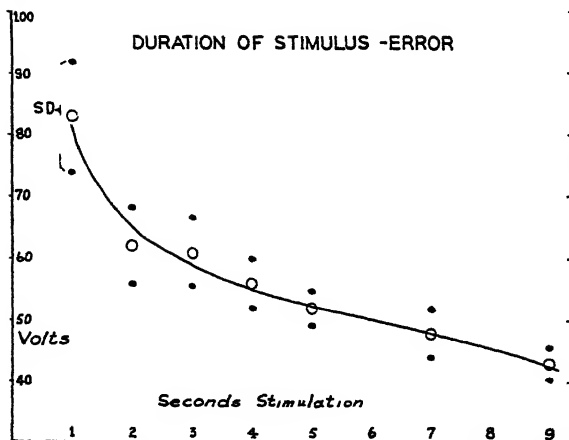


FIGURE 1. Effect of increasing the duration of stimulation on reactive threshold of the guinea pig. Note that the standard deviation of the response becomes smallest with a stimulus of five seconds duration.

aspirin¹³ and other weak analgesics by this method.¹⁶ Chemicals which may give false positive analgesia by this method are those that are curare-like or myanesin-like drugs, and drugs which produce hypothermia, methemoglobinemia, hypoxia, or marked elevation of blood pressure. Drugs which lower the pain threshold are acetylcholine, *L*-arginine, *L*-ornithine, aranthol,¹⁷ convulsants and non-anesthetic doses of the barbiturates. Since Demerol and codeine are both convulsant in the higher dosage range, we find a characteristic decreased effect after the thirty-minute period when subconvulsive doses are used (FIGURES 2 and 3).

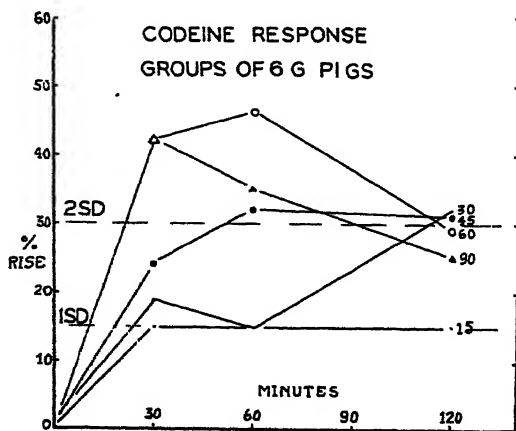


FIGURE 2. Effect of increasing dosage of subcutaneously administered codeine phosphate in groups of six guinea pigs. Note that higher subconvulsant dose of 90 mg. kg. produces an apparent decrease in the analgesic effect.

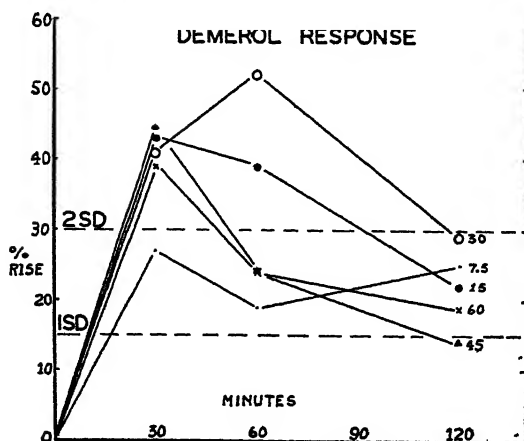


FIGURE 3. Effect of increasing dosage of subcutaneously administered Demerol HCl. Note that the higher subconvulsant doses of 45 and 60 mg./kg. produce an apparent decrease in the analgesic effect.

Comparison of Three Methods in Man. Since previous studies have dealt with a single type of pain (usually superficial pain), we initiated a program early last fall which was designed to compare the two thresholds of the tooth-pulp method with the superficial pain threshold as elicited by radiant heat in the pad of the fingers, after the hand has been immersed for one minute in a constant temperature bath maintained at 40° C. This method provides a constant skin temperature and avoids the production of forehead blisters and scars which invariably follow analgesic testing when potent analgesics are assayed by the Hardy-Wolff method. A third threshold was also assayed when H. L. Williams of our laboratory noted that, if the fingernail bed is subjected to radiant heat, a marked ache of an all-or-none quality was produced. In a series of thirteen trials using the 3rd and 4th fingers of two subjects, the range between the pain threshold and the reactive threshold was found to be $30.2 \pm \text{S.D. } 1.7$ watts for the nail bed and $63.8 \pm \text{S.D. } 11.4$ watts for the finger pad. Waterproof dull black lacquer (Eastman Kodak Co.) is used to coat both the nail and the pad. Although the nail must be thicker than the usual keratin layer on the pad, the nail threshold in watts is characteristically about 60–70 per cent of that of the corresponding pad. The index finger has the highest threshold, the thumb next, while the little finger is the most sensitive. The third and fourth fingers are best for the testing of analgesic drugs. Fortunately, we discovered early in our studies that in all fingers (with the possible exception of the little finger) the pain threshold of the pad is in balance with that of the nail bed. This may be demonstrated by injecting a drop of 0.5 per cent procaine in the finger pad, which usually makes the nail bed threshold decrease (become more sensitive). We have also tested nail bed and pad of the middle finger after 0.5 per cent procaine has been infiltrated around the median

nerve. This, characteristically, results in a rise in the pad threshold and a lowering of the nail bed threshold (FIGURE 4). If 0.75 per cent procaine

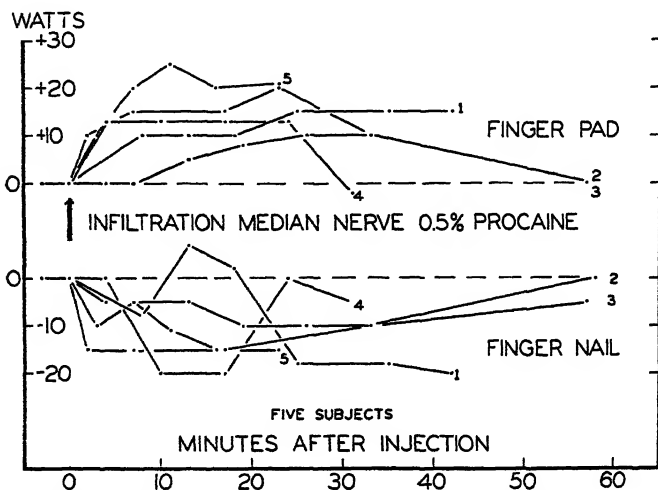


FIGURE 4. Effect of infiltration of the median nerve on the finger pad and nail-bed pain thresholds of the middle (3rd) finger of five subjects. Note that the pad threshold is raised while the nail bed threshold is lowered.

is infiltrated around the median nerve, both thresholds will rise, but as recovery occurs the nail bed will suddenly drop to well below the initial control thresholds while the pad threshold will remain elevated. Thus, in so far as the middle finger is concerned, the sensory nerves to the pad and nail bed are both contained in the median nerve, but the nail bed nerves may be more centrally located in the nerve trunk or, perhaps, may be more resistant to local anesthetic action. Central cortical summation, balance mechanisms in the internuncial pool of the spinal cord, or local axonic networks (nocifensor fibers?) between the fingernail bed and the pad may explain this phenomenon, which has also been observed during analgesic testing; that is, sudden rises in pad thresholds are apt to be accompanied by an increased sensitivity of the nail bed. However, with analgesic drugs the changes are small, and the mean of the data thus serves to obliterate these minor individual fluctuations.

Since tooth pain is aching in character and quite different from the pinch or tingle of beginning skin or superficial pain, we have been interested in determining the possible type of pain which may predominate in the nail bed to produce the all-or-none ache. Heavily myelinated nerve fibers such as epicritic or supain fibers have been reported to be more susceptible to hypoxia produced by ischemia. Accordingly, we have determined nail-bed and pad pain thresholds after ligaturing the arm with a blood pressure cuff inflated to 250 mm. Hg for periods of 15 to 40 minutes. In thirteen experiments we have found the pad when tested

after warming to 40° C. to become more sensitive in five experiments, analgesic in 4, and no change in 4. The fingernail bed in these same experiments became more sensitive in 5 experiments, analgesic in 1, and no change was noted in 7 experiments. These data are thus equivocal and cast doubt on the validity of previous ischemia experiments, where no attempt was made to test the ligatured arm at a constant temperature. We have been able to differentiate quantitatively between nail-bed and pad pains by the differential rise in thresholds when the finger is tested at less than 40° C. (TABLE 2). As the hand temperature is lowered below

TABLE 2
EFFECT OF TEMPERATURE OF THE HAND
ON RADIANT HEAT PAIN THRESHOLDS OF THE FINGER

(Mean of two subjects hand immersed for one minute before testing—arm not ligatured.)

	<i>Fingernail bed</i>	<i>Finger pad</i>
45° C.	+20%	-10%
40° C.	0%	0%
35° C.	+25%	+30%
30° C.	+40%	+27%
20° C.	+70%	+33%
10° C.	+110%	+66%

35° C., the rise in nail-bed pain threshold rapidly becomes twice the pad pain threshold rise. Extreme or maximal tolerated heat (46–47° C.) does not consistently differentiate the two thresholds. Since no correlation was found between duration of ligation and trends in the nail-bed and pad thresholds, these results indicate the need for raising the ligatured hand to a constant temperature before testing with radiant heat, since Wolff *et al.*¹⁸ have reported that the back of the hand of the ligatured arm (one or possibly two subjects!) has the same degree of analgesia as the forehead when the pain of the ligatured arm raises the forehead threshold 30 to 35 per cent. In our experience, this rise in the pain threshold of the finger pad or nail bed can be obtained with the radiant heat method when skin temperature decreases only 3 to 5° C. This degree of lowering of hand temperature will occur when the arm is ligatured. The relief of pain by hypothermia may explain the fact that patients with severe causalgia invariably carry a moistened sponge to apply cooling water to the painful extremity.

Tetraethylammonium (TEA) chloride in a dose of 500 mg. was next given intravenously to a group of five subjects on the working hypothesis that, since this drug is effective in causalgia, it might raise nail-bed pain thresholds. To our surprise, the initial doses of TEA raise only the supain thresholds of the finger pad and leave tooth pain and, to a lesser extent, nail-bed pain thresholds remarkably constant in spite of marked subjective symptoms such as diplopia, cycloplegia, and tingling of the extremities (TABLE 3). Repeated daily doses of TEA, as used in causalgia, will produce a rise in the thresholds of nail bed and tooth. The effect of

TABLE 3

EFFECT OF TETRAETHYLAMMONIUM CHLORIDE (500 MG.) ON VARIOUS PAIN THRESHOLDS

(All doses injected in 8 to 10 minutes. A consistent rise is seen in the finger-pad pain threshold in the 12-to-30-minute period. Index (2nd) finger used in all subjects.)

Subj. C.P.	Controls		4	10	12	15	20	30	35	40	Minutes
Nail Bed	85	85	90	90	90	90	85	90	85	85	Watts
Finger pad	110	110	120	135	130	135	125	120	120	110	Watts
Tooth I	0.6	0.5	0.5	0.5	0.5	0.6	—	—	0.5	—	Volts
Tooth II	2.6	2.7	2.2	2.6	2.8	2.8	—	—	2.8	—	Volts

Subj. R.S.	Controls		1	5	10	15	25	32	35	40	Minutes
Nail bed	90	90	95	105	90	90	85	90	92	85	Watts
Finger pad	120	120	120	145	145	135	140	125	130	125	Watts
Tooth I	1.1	1.1	1.0	1.1	1.1	—	1.0	1.1	1.1	1.2	Volts
Tooth II	1.6	1.6	1.4	1.8	1.6	—	1.6	1.6	1.5	1.7	Volts

Subj. J.L.	Controls		1	5	9	14	19	30	40	60	Minutes
Nail bed	60	65	65	55	60	65	60	65	60	60	Watts
Finger pad	82	82	90	95	95	95	100	105	97	82	Watts
Tooth I	0.5	0.5	0.5	0.5	0.5	0.5	0.6	0.6	0.5	0.5	Volts
Tooth II	1.2	1.1	1.4	1.2	1.2	1.1	1.1	1.1	1.0	1.0	Volts

TEA on supain thresholds cannot be central since deepain is unaffected.

Ergotamine tartrate does not raise supain thresholds,¹⁹ but does relieve migraine headache, perhaps by local or peripheral vasoconstriction. If nail-bed pain is due to local vascular dilatation or spasm, ergotamine tartrate might decrease or increase this pain threshold. Accordingly, two subjects were given subcutaneously $\frac{1}{2}$ mg. of ergotamine tartrate and three subjects were given subcutaneously 1 mg. of DHE-45 under blind test conditions. Neither of these drugs changes supain, deepain, or nail-bed pain thresholds in these normal subjects (TABLE 4).

TABLE 4

MEAN EFFECT OF ERGOTAMINE TARTRATE

(2 SUBJECTS) AND DIHYDROERGOTAMINE (3 SUBJECTS) ON PAIN THRESHOLDS

(The middle (3rd) finger used in all subjects. A possible decrease in nail-bed pain threshold and Tooth II threshold may obtain.)

	Controls		20	40	60	80	100	120	Minutes
Nail bed	65.5	67.0	62.0	62.0	64.5	65.5	66.0	60.5	Watts
Per cent change	—	—	-6.6	-6.6	-2.6	-1.1	-0.4	-8.7	
Finger pad	109.5	112	110.2	106.5	110.4	106.4	111.6	111.0	Watts
Per cent change	—	—	-0.5	-3.8	0.3	-3.9	0.8	0.2	
Tooth I	0.49	0.51	0.48	0.48	0.54	0.50	0.50	0.45	Volts
Per cent change	—	—	-4.0	-4.0	8.0	0.0	0.8	-10	
Tooth II	1.93	1.91	1.86	1.74	1.84	1.8	1.75	1.70	Volts
Per cent change	—	—	-3.1	-9.4	4.2	-6.2	-8.8	-11.5	

A slight increase in sensitivity of deepain and nail-bed pain occurs. We were, however, impressed with ergotamine tartrate as a possible control medication for the study of opiate analgesia. The symptoms of sweating

and nausea are sufficiently similar to opiate side effects that this is the first placebo medication which our trained subjects have not been able to identify as a placebo medication within 15 or 20 minutes. The side actions are more rapid in onset than the side actions of potent analgesics, so that even this medication can be identified if used a second time in the same subject.

Stellate ganglionic block with 1 per cent procaine will produce a prompt rise in the nail-bed pain threshold of the middle finger without a corresponding rise in supain thresholds of the finger pad. If the somatic nerves are later affected by the procaine, a sudden drop in nail-bed pain threshold may occur, indicating again the reciprocal relationship of the pad and fingernail bed (FIGURE 5). Such successful stellate block has

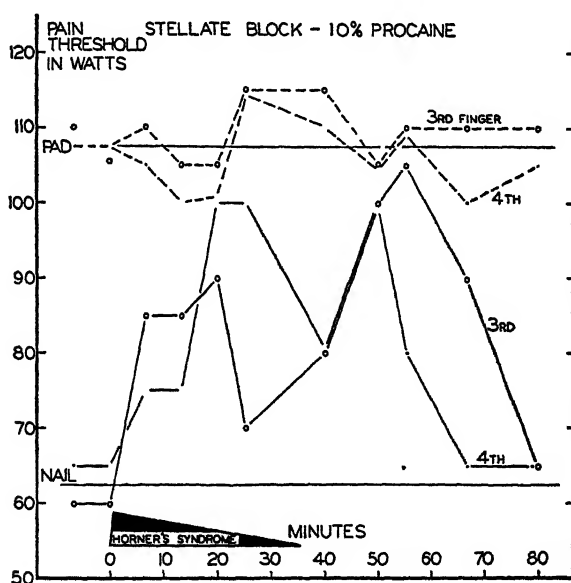


FIGURE 5. Effect of right stellate ganglionic block with 1 per cent procaine on the 3rd and 4th finger pain thresholds. Note again the reciprocal relationship of nail-bed pain to finger-pad pain during the 20-to-40-minute period when the procaine apparently obtunded somatic afferent nerves.

now been obtained in three trials on two subjects. The rises in nail-bed threshold have been 91 per cent, 50 per cent, and 54 per cent. To elucidate further the possible mechanism of nail-bed pain, a painful heat stimulus has been applied to the scarified fingernail when viewed with the highest power of a dissecting microscope. Under these conditions, the capillaries do not contract with pain, but the arterioles cannot be seen. From our present data, this pain could thus be either efferent or afferent sympathetic in origin. Since it is obtunded by stellate block we have termed it "sympain" and will attempt further experiments to determine whether it is efferent or afferent. In this regard, the fact that sympain

is almost entirely unaffected by TEA argues against nail-bed pain being due to efferent sympathetic action. Undoubtedly, the nail bed may be subserved in addition by protopathic and perhaps epicritic nerve fibers. We only wish to claim from these data that the nail bed has a strong sympathetic component.

Recalling that sympain is in balance with supain, the data obtained thus far with potent analgesic drugs can now be studied critically. The three drugs we have chosen for study in 5 to 7 subjects are heroin (2 mg.—TABLE 5), Dilaudid (2 mg.—TABLE 6), and *l*-methadon (5 mg.—TABLE 7).

TABLE 5

ANALGESIA PRODUCED BY THE SUBCUTANEOUS INJECTION OF 2 MG.
HEROIN HCL IN SEVEN NORMAL SUBJECTS

(Note the lack of effect on nail-bed pain and the marked effect on tooth pain.)

	<i>Controls</i>		20	40	60	80	100	<i>Minutes</i>
Nail bed	91.9	92.3	100	92.3	90.7	89.6	92.1	Watts
Per cent change	—	—	8.6	0.2	-1.5	-2.7	0.0	
Finger pad	117.9	117.8	124.3	129.3	127.9	128.6	127.9	Watts
Per cent change	—	—	5.8	10.0	8.9	9.4	8.9	
Tooth I	0.72	0.75	0.95	1.0	0.93	0.84	0.88	Volts
Per cent change	—	—	30	36	27	15	20	
Tooth II	1.70	1.73	2.4	2.8	3.0	2.8	2.7	Volts
Per cent change	—	—	41	64	73	64	60	

In two-thirds of the actual experiments, these drugs were given in blind-tests. Heroin was chosen for its short duration of action and its maximal addiction liability. Dilaudid was chosen because of its potency at a low dose and because it is, in the monkey, less addicting than morphine in a dosage ratio of 1 to 5. *l*-Methadone was included since studies in the monkey and man indicate that it is the least addicting of the potent analgesic drugs and, on withdrawal, produces a slow metabolic storm rather than the immediate autonomic storm which is seen on opiate withdrawal.

It is evident, from the individual drug tables and the summary (TABLE 8), that these three drugs vary considerably in their effect on sympain, supain, and deepain thresholds. Thus, Dilaudid has 15 times the effect of heroin on sympain, and methadone is characterized by a broad spectrum which produces an equal effect on all types of pain. The abstinence syndrome has been characterized as an "autonomic release phenomenon" which might now be interpreted as an overdevelopment of sympain, due to the fact that heroin markedly obtunds deepain and has little or no effect on sympain. Thus, repeated doses may allow the sympain component to overdevelop and produce symptoms of autonomic over-activity when a large heroin dosage is abruptly stopped. A drug such as *l*-methadone, which raises all pain thresholds uniformly, is less apt to be followed by an abstinence syndrome which is characterized by autonomic symptoms. This working hypothesis for opiate addiction is

TABLE 6

ANALGESIA PRODUCED BY THE SUBCUTANEOUS INJECTION OF 2 MG. DILAUDID HCl IN SEVEN NORMAL SUBJECTS
(Note that dilaudid has a greater effect on nail-bed pain than heroin and note the marked rise in tooth thresholds.)

	<i>Controls</i>		20	40	60	80	100	120	140	160	<i>Minutes</i>
Nail bed	71.4	71.7	78.6	82.9	80.7	83.6	87.0	77.5	78.6	73.6	Watts
Per cent change	—	—	8.4	16	13	17	23	8.4	9.8	2.8	
Finger pad	112.9	114.3	117.1	121.4	127.1	132.1	126.4	125.0	122.1	119.6	Watts
Per cent change	—	—	3.1	6.9	11.9	17.1	11.2	10.0	8.2	5.3	
Tooth I	0.61	0.61	0.74	0.78	0.93	0.85	0.88	0.82	0.75	0.67	Volts
Per cent change	—	—	21	28	52	39	44	34	23	10	
Tooth II	1.93	1.93	2.3	2.8	3.3	3.3	3.3	2.9	2.8	2.4	Volts
Per cent change	—	—	19	44	71	71	71	52	42	22	

TABLE 7

ANALGESIA PRODUCED BY THE SUBCUTANEOUS INJECTIONS OF 5 MG. OF *l*-METHADONE HCL
IN FIVE NORMAL SUBJECTS

(Note uniformity of effect on all pain thresholds.)

	Controls	20	40	60	80	100	120	140	160	Minutes
Nail bed	69.4	72.0	77.0	82.0	86.4	89.8	82.4	87	87	Watts
Percent change—	—	—	9	16	22	27	17	23	23	
Finger pad	98	100.6	111	120	113	115	113	115	119	Watts
Percent change—	—	—	12	21	14	16	14	16	20	
Tooth I	0.50	0.54	0.63	0.74	0.66	0.65	0.61	0.61	0.58	Volts
Percent change—	—	—	21	42	27	25	17	17	12	
Tooth II	2.06	2.08	2.6	2.7	2.6	2.6	2.7	2.5	2.3	Volts
Percent change—	—	—	26	31	27	28	31	20	12	

TABLE 8

SUMMARY OF MEAN PER CENT RISE IN PAIN THRESHOLDS WITH HEROIN (2 MG.),
DILAUID (2 MG.), AND *l*-METHADONE (5 MG.).

Drug	Nail bed	Finger pad	Tooth I	Tooth II
Heroin (100 minutes)	1%	8%	26%	60%
Dilauid (160 minutes)	15%	9%	26%	50%
<i>l</i> -Methadone (160 minutes)	20%	17%	21%	24%

being further tested by studies of the analgesic effect of desomorphine and *l*-isomethadone. TABLE 9 is a final summary of the procedures which differentiate these three types of pain.

TABLE 9

SUMMARY OF METHODS AND DRUGS USED TO DIFFERENTIATE
THREE TYPES OF PAIN IN MAN

	<i>Sympain</i>	<i>Supain</i>	<i>Deepain</i>
Ischemia	±	±	Not tested
Hypothermia	2+	+	Not tested
Quality	Ache	Tingle	Ache
	(all-or-none)		
Procaine 0.5% in median nerve	Decreased threshold	Increased threshold	Not tested
TEA (500 mg.)	±	2+	0
Ergotamine	—?	0	—?
Stellate block	4+	0	0
Heroin	0	+	4+
Dilauid	2+	+	4+
<i>l</i> -Methadone	2+	2+	2+

Summary

Two statistically evaluated methods are available for the determination of the reactive threshold of animals. These are: (1) The use of radiant heat with a constant stimulus and variable time, and (2) the use of radiant heat with a variable stimulus at a constant time interval. The use of contact heat might be improved by the use of a sloping hot plate

in an apparatus analogous to the sloping screen method for insulin assay in the mouse.

Three accurate methods are available for the assay of pain thresholds in man. These are: (1) The use of radiant heat with a variable stimulus and constant time; (2) the use of electrical stimulation of the tooth pulp which provides the minimal and reactive thresholds; and (3) the use of von Frey hairs on sensitive areas of the face.

Three different types of pain may exist in man. In order to encourage sub-division of these pains by clinicians, the following terms are suggested: for bright, fast, epicritic, superficial pain the term "supain" is suggested. For dull, aching, slow, protopathic pain the term "deepain" is applied. Nail-bed pain has been shown to have a strong sympathetic component and hence the term "sympain" is suggested. By comparing the analgesic action of heroin and *l*-methadone in clinical patients and the use of sympathetic blocks, further instances of sympain may be found.

Since heroin raises deepain thresholds but has little effect on sympain, and since *l*-methadone raises all pain thresholds uniformly, a working hypothesis for the possible etiology of the abstinence syndrome has been formulated which is predicated on the probability that a central or local balance obtains between various types of pains. This balance is demonstrated for the supain and sympain of the finger.

Bibliography

1. GOETZL, F. R. 1946. *Permanente Found. Med. Bull.* 4: 49.
2. SEEVERS, M. H., & C. C. PFEIFFER. 1936. *J. Pharm. & Exp. Therap.* 56: 166.
3. SEEVERS, M. H., J. H. BENNETT, H. W. POHLE, & E. W. REINARDY. 1937. *J. Pharm. & Exp. Therap.* 59: 291.
4. MACHT, D. I., & M. B. MACHT. 1940. *J. Am. Pharm. Assoc. Sci. Ed.* 29: 193.
5. LANIER, L. H. 1943. *Science* 97: 49.
6. SCHUMACHER, G. A., H. GOODELL, J. D. HARDY, & H. G. WOLFF. 1940. *Science* 92: 110.
7. KOLL, W., & H. REFFERT. 1938. *Arch. Exp. Path. & Pharmacol.* 190: 176.
8. IVY, A. C., D. Y. BURRILL, & F. R. GOETZL. 1943. *Quart. Bull. Northwest Univ. Med. Sch.* 17: 280.
9. HARRIS, S. C. 1948. *Northwest Univ. Dental School* (Personal communication).
10. SONNENSCHIEIN, R. R., R. JAMISON, L. LOVSETH, W. CASSELS, & A. C. IVY. 1948. *Fed. Proc.* 8: 118.
11. WOOLFE, G., & A. D. MACDONALD. 1944. *J. Pharm. & Exp. Therap.* 80: 300.
12. D'AMOUR, F. E., & D. L. SMITH. 1941. *J. Pharm. & Exp. Therap.* 72: 74.
13. WINDER, C. V., C. C. PFEIFFER, & G. L. MAISON. 1946. *Arch. Int. Pharmacol.* 72: 329.
14. HARDY, J. D., H. G. WOLFF, & H. GOODELL. 1940. *J. Clin. Invest.* 19: 649.
15. ANDREWS, H. L., & W. WORKMAN. 1941. *J. Pharm. & Exp. Therap.* 73: 99.
16. PFEIFFER, C. C. 1942. Unpublished experiments.
17. PFEIFFER, C. C., J. SANTOS-MARTINEZ, & T. R. SHERROD. 1948. *Fed. Proc.* 7: 248.
18. WOLFF, H. G., J. D. HARDY, & H. GOODELL. 1940. *J. Clin. Invest.* 19: 659.
19. WOLFF, H. G., J. D. HARDY, & H. GOODELL. 1941. *J. Clin. Invest.* 20: 63.

A CRITIQUE OF ANALGESIC TESTING METHODS

By LLOYD C. MILLER*

Sterling-Winthrop Research Institute, Rensselaer, N. Y.

IN 1943, Goetzl, Burrill, and Ivy¹ reviewed the literature on methods used up to that time to evaluate pain intensity and the analgesic effects of drugs. No method new in principle has appeared since then, so that their analysis may well be taken as a point of departure for this discussion. These authors set up four criteria which the ideal method should meet in yielding quantitative data on the extent to which an analgesic obtunds pain.

They considered that the ideal method should permit quantitative determination of threshold values of the selected stimulus or means of inducing pain. It is a foregone conclusion that laboratory methods for algometry must be designed to prevent or ameliorate pain induced artificially. From a pharmacological standpoint, this is in keeping with principles well established in other fields. Thus, antispasmodics are tested against spasms induced in smooth muscle of various kinds; likewise, antihistaminics are rated according to their capacities for alleviating or preventing effects elicited by histamine. Stimuli resulting in pain closely approach those which produce permanent damage. All authors agree that for heat, at least, frank burns result from increasing the threshold pain stimulus only three- or four-fold so that it is of great importance to determine accurately the threshold or just-effective stimulus. The reasons for this are obvious in the case of human subjects. With animals, damage is to be avoided to permit repeating observations on the same animal following both the test drug and a reference analgesic, as in a cross-over test.

The ideal method should discriminate well between graded doses of an analgesic in modifying the responses to a standard pain stimulus. Thus, it must be a biological assay in the fullest sense of the term. It is in this respect that many methods have failed.

The ideal method should be universally applicable to both man and experimental laboratory animals. Because pain is a subjective phenomenon, more difficulty arises on this point with animals than with human subjects, a situation which is the reverse of that usually encountered in pharmacological studies.

Goetzl and his associates also deemed that the ideal method should show quantitatively the respective effects of the analgesic against different qualities of pain, where such exist. Since an effort has been made in devising most methods to select a stimulus giving a single kind of pain, no one method will meet this criterion.

* The assistance of Mr. Calvin Zippin in the calculations and in drawing the graphs is gratefully acknowledged.

Finally, for reasons peculiar to the pharmaceutical industry, there should be added to the four requirements listed above, the specification that the ideal method should boast of sensitivity. This feature is important to chemists and pharmacologists alike, for valuable leads to analgesic activity in new types of compounds may well be overlooked if the test method is not capable of revealing low grades of effect. Unfortunately, no method measures up well in this respect.

If this discussion is restricted to those methods used most generally in evaluating the newer synthetic analgesics, it will be possible to show how these procedures measure up to the ideal. While the relatively crude Haffner technique² was sufficient in Schaumann's hands to show in classic fashion the value of Demerol and related compounds in 1939,³ Woolfe and Macdonald,⁴ in 1944, considered that the potency of Demerol was still inadequately evaluated. Still later, in 1947, six laboratories collaborated in comparing Demerol and three other analgesics with morphine, with very heterogeneous results⁵; for example, there was a nine-fold variation in the observed potencies of Demerol expressed in terms of morphine. It would appear worth-while to explore the causes for this failure on the part of numerous pharmacologists.

Goetzl, Burrill, and Ivy concluded that the method most closely approaching the ideal would apply a stimulus, consisting of a sharply defined burst of electrical current, to affect the specific pain receptors of the tooth pulp of either man or laboratory animals. Such a method had been described by Koll and Reffert in 1938,⁶ who used a condenser discharge stimulator with highly consistent results on dogs. Goetzl and his associates used a stimulator designed to provide peaks of induced alternating current. However, while the characteristics of the apparatus were described fairly clearly, they failed to give details of dimensions, etc., sufficient to permit others to reproduce it. Further, details of the performance of their method are lacking, so that there is no basis for forming an opinion as to its reproducibility and discriminatory power. In our laboratory, a serious attempt was made to develop this technique using dogs, but it was found impossible to reduce the responses of the animals to anything like a constant base line. Except that this technique is bound to be relatively slow, the success of Koll and Reffert with it recommends further trial. Finally, there have been no reports of the use of this method in studies of recently discovered analgesics, so that further consideration of it is beyond the scope of this paper.

The Hardy-Wolff-Goodell Method. Of the methods proposed and used in the last decade, it is safe to say that the Hardy-Wolff-Goodell procedure⁷ has gained the widest acceptance. It was described, in 1940, by a team of three workers who doubtless will be remembered as having experienced more self-inflicted pain than any other trio in the history of experimental medicine.

Almost everyone interested in algesimetry is familiar with the tech-

nical details of the Hardy-Wolff-Goodell method. Briefly, it consists of allowing the light from a 500-watt projection lamp to focus for exactly three seconds on the subject's forehead, the exposed area being blackened with ink. The lamp operates continuously on reduced voltage and the pathway of its beam is opened and interrupted by a shutter controlled by an electrical timer. The only variable is the current applied to the lamp filament so that, by increasing the current in successive exposures, the intensity of stimulus is raised finally to the point at which the subject feels a brief, sharp "jab" of pain just at the end of the three-second exposure period. Originally, the time between successive trials was only 30 seconds, but later it was increased to a full minute.

By nature, pharmacologists have a penchant for tinkering and seldom resist an opportunity to add a personal twist or variation to each new method or technique they take up. It is interesting, therefore, that the Hardy-Wolff-Goodell method remains today almost exactly as originally proposed. Indeed, apparatus specially designed for the procedure is available commercially* which, in itself, is a good indication that the technique has become stabilized. Through the kindness of Drs. Goodell and Wolff, it has been possible to compare data obtained nearly 10 years ago with the original apparatus, with some recorded by Dr. Gross, of the State University of Iowa, using the commercial unit.

The older data were those obtained in determining the analgesic potency of codeine⁸ and represent seven experiments in which the dosage of codeine phosphate was varied from 15 to 120 mg., injected intramuscularly. Usually, three readings were made to establish the normal threshold prior to injection of the drug, although this was not always the case. Because of these minor exceptions, some liberties were taken with the data on the grounds of convenience.

An analysis of variance of the normal thresholds in the five experiments in which three normal threshold readings were made, shows that the variation between these three values, as well as that between the three subjects, was quite negligible. The variation between days, however, was enormous, being far greater than would be expected once in a thousand times through chance. This indicates an obscure variation in the apparatus or environment which resulted in apparent differences in thresholds on different days. Nevertheless, the standard deviation of a single observation was only ± 17 millicalories/second/square centimeter, or about 5 per cent of the average. As low as this error seems to be, it is large in comparison with what may be achieved.

Similar analysis of the Iowa data shows a truly remarkable consistency. The experiment involved the subcutaneous injection of a new analgesic, Nu-1779,⁹ into four subjects, each of whom received 5, 10, or 15 mg. on separate days. The normal pain thresholds were determined in two trials after an initial 10-minute rest period, so that there are six nor-

* The Experimental Engineering Corporation, Bergenfield, N. J. has developed a commercial model of the Hardy-Wolff-Goodell apparatus.

mal values on each subject, or 12 pairs in all. These 24 readings all lay between 228 and 235 millical./sec./cm.² This amazing consistency is reflected in the analysis of variance which shows that the variation between the four subjects was practically nil, that between the three experimental days was even less, while discrepancy between pairs of readings on the same day was greater, but still far from significant. As Dr. Gross intimated to me in correspondence, he is actually concerned over the fact that the data appear incredibly good.

The Effect of Training on the Normal Pain Threshold. These subjects had been trained. With respect to the value of training, Dr. Gross sent interesting data obtained in the course of "breaking in" four "green" subjects on the Hardy-Wolff-Goodell apparatus. This was a well-balanced experiment, in that the pain threshold of each subject was determined four times at regular intervals on each of four days. These are essentially "sham" experiments except that no injections were made; the subjects were unaware of the actual amount of heat stimulus being applied. The individual readings in millical./sec./cm.² are shown in TABLE 1. The averages are given in the parentheses to the nearest whole number. To bring out the day-to-day improvement, the sums of the

TABLE 1
EFFECT OF TRAINING ON PAIN THRESHOLD RECOGNITION

(Hardy-Wolff-Goodell Method (Gross); normal thresholds in millical./sec./cm.²)

Subj. No.	Day 1	Day 2	Day 3	Day 4	Subj. totals*
1	250	240	230	235	
	240 (243)	230 (234)	235 (232)	230 (232)	
	245	230	222	232	
	238	235	230	230	
	53*	15	7	7	
2	235	240	232	230	
	240 (238)	230 (235)	230 (232)	230 (231)	
	245	235	235	223	
	230	235	230	230	
	30	20	7	3	
3	235	230	235	235	
	240 (236)	235 (235)	235 (233)	230 (232)	
	240	233	230	233	
	230	240	232	230	
	25	18	12	8	
4	230	235	235	230	
	230 (234)	235 (235)	230 (232)	235 (231)	
	240	240	230	230	
	235	240	232	230	
	15	20	7	5	
Day totals*	123 (238)	73 (235)	33 (232)	23 (231)	47

* Sums of deviations from 230.

deviations from 230, the lowest observed threshold, are given. This is still more evident in the progressive decrease in the day totals from the first to the fourth days.

A brief glance at the average values shows that the subjects had settled down to practically the same threshold value by the second day and, while they continued to agree among themselves, the actual threshold decreased somewhat from the previous day's run on both the third and fourth days.

To determine which of these differences was significant, an analysis of variance was run. The analysis (TABLE 2) showed that the four subjects

TABLE 2
ANALYSIS OF EFFECT OF TRAINING

Source of variation	Degrees of freedom	Mean square	Measure of significance	
			F	P
Subjects	3	13.04	.77	
Runs	3	27.85	1.64	>.05*
Days	3	129.17	7.61	<.01†
Error	54	16.97		

* for $P=.05$, $F=2.77$. † for $P=.01$, $F=4.16$. F and P have the conventional meaning, so that $P=.05$ indicates a probability of 5 chances in 100, etc.

reacted very much alike, the variance attributable to differences between them being less than that ascribable to "error." The variance between runs on the same day was also quite insignificant, although greater than that between subjects. From day to day, however, the variation was highly significant, the odds being less than one in a hundred that such discrepancies would arise by normal chance variation. Because of the marked downward trend, the conclusion is justified that, with training, each subject comes to recognize the intensity of heat stimulus marking his threshold and that the threshold is very nearly the same for all subjects. This view, of course, has long been advocated by Hardy and his associates.¹⁰

Denton, Straus, and Beecher¹¹ recently reported that untrained subjects are unsuitable for evaluating analgesics by this method. Thus, it is possible that a lack of training may have been responsible for the failures on human subjects reported recently from two British laboratories.^{12, 13}

In a study of the Hardy-Wolff-Goodell procedure applied to guinea pigs, Winder has explored the effect of practically all imaginable variables possible. An adequate discussion of his findings, published in a series of four papers,¹⁴⁻¹⁷ would overtax the limits on available space.

Winder chose the graded intensity procedure, approaching the threshold by "interval-splitting," gradually narrowing the difference in the intensity of 4-second exposures to finally locate the threshold within five watts of power input on the lamp. This is a relatively slow and tedious procedure since it takes 10 minutes to find the thresholds on two ani-

mals tested alternately at successive half-minute intervals. These animals were used but once, so that, except for repeated exposures within the experiment, no training was provided. Repeated threshold endpoints on the same animal agreed with each other better than those obtained on different animals run in parallel; thus, the final procedure was designed so that each animal served as its own control. To correct for a slight but significant downward trend in the threshold during the experiment, sham-treated animals, given only the drug vehicle, are run in parallel.

Bonnycastle¹⁸ finds it necessary even to train rats prior to use. Thorp¹² also states that new rats must be trained "to discriminate between stimuli of different intensity before using them for the test." The question arises as to where the greatest amount of intelligence is required—in the rats or in the operators! In our experience with a similar procedure, using rats, no greater variation has been noted on the first day of use than in subsequent trials.

Interpretation of the Data. In spite of the rather extensive use of the Hardy-Wolff-Goodell method on both man and laboratory animals, no good way of handling the data has been worked out. It is customary to make observations at intervals of 15 to 30 minutes throughout the experiment, so that the data consist of four to 10 observations on each subject for each dose. The peak effect is usually seen between 20 and 60 minutes. The peak values, calculated as the percentage increase over the pre-treatment threshold value, are usually averaged for each dosage. This results in using only a part of the data, as has been done by Thorp¹² with the result shown in FIGURE 1 for morphine given to rats subcuta-

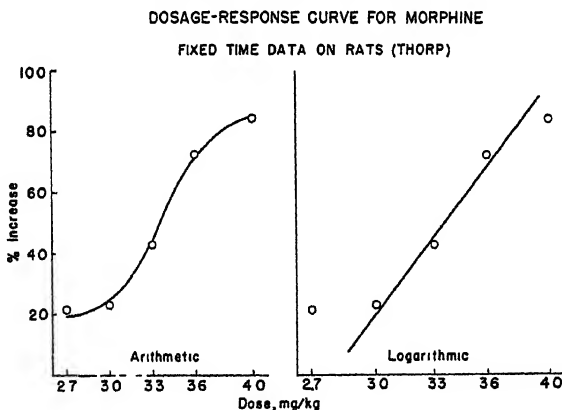


FIGURE 1.

neously. A beautiful sigmoid curve results when the per cent increases at 20 minutes are plotted against dose either in the original arithmetic units or in logarithmic units. Although this is one of the best dosage-

response curves yet published for morphine, Thorp did some violence to the data by drawing a straight line through what he termed the "central portion." Ignoring the obvious curvature, he even went to the trouble of calculating the equation of this "linear" part.

As evidence of the reliability of this treatment, he assayed two unknown solutions of morphine with interesting results. With solution B, a mean percentage increase of 63.1 per cent was observed from which the concentration was estimated to be 0.35 mg./cc. against the actual strength of 0.36 mg./cc. With solution A, he obtained a 20.6 per cent increase in threshold, from which fact he estimated that the concentration was 0.29 mg./cc.; its actual strength was 0.30 mg./cc. The agreement on solution B between found and actual concentration is not surprising, but it was sheer luck to come so close in determining the strength of solution A, when the dosage-response curve in the neighborhood of 20 per cent increase in pain threshold is in the flat part of the sigmoid curve, as may be seen from FIGURE 1.

Winder¹⁷ plotted the logarithm of the maximum threshold against log dose for data he obtained on morphine, Demerol, and aspirin. Here again, only the data obtained at the time of peak effect have been utilized.

Some means is needed for taking into account all the data obtained. It seems likely that the neglected values might do more than merely bolster the peak readings. There is some indication that the discrimination of the method might well be greater at other points.

To data obtained on dogs by Gross in the above-mentioned collaborative study, Bliss⁶ applied the statistical technique known as "discriminant function analysis."¹⁸ This is a relatively recent development, the basic principle of which may be deduced from FIGURE 2, which presents further unpublished data obtained on human volunteers by Dr. Gross with Nu-1779.⁹ It is assumed that each observation contributes some information towards "discriminating" between the different dosage levels used. As shown in the graph, the individual readings obtained at the successive times of observations may be designated as y_1, y_2 , etc. The object of the analysis is to determine those coefficients for the various values of y in an algebraic equation relating dose to effect which maximizes the discrimination between the various drug levels. Unfortunately, this calculation is far from simple. It would be especially tedious for the data shown here with 10 different values of y and has not been carried out. Needless to say, a simpler method of determining the coefficients must be evolved if this method is to become popular.

Of greater interest, in this graph, is the visual evidence on the extent to which the effect of doubling the dose is reflected in the response of the subject. While this is one of the better examples obtained in this experiment, 100 per cent increments in dose can readily be distinguished nevertheless.

The point on which the Hardy-Wolff-Goodell procedure is the most

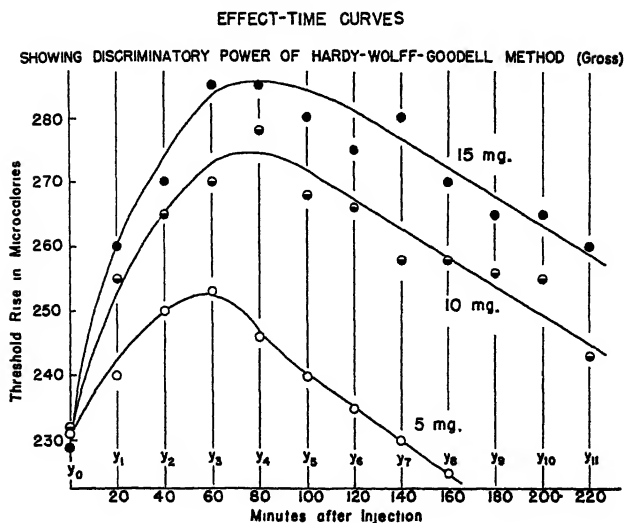


FIGURE 2.

vulnerable is the adverse influence that emotional and psychological distractions have upon the pain threshold. Dr. Wolff has himself emphasized this point.²⁰ He has observed that, in a susceptible subject, the effect of 0.3 gm. of aspirin may be either duplicated or completely masked. On the other hand, subjects not amenable to suggestion show no demonstrable effect upon receiving placebos. This suggests rather forcibly the desirability of including placebos as an integral part of the design of critical experiments.

D'Amour-Smith Method. Although the Hardy-Wolff-Goodell procedure has rather successfully resisted efforts to "improve" it, there is an important group of publications based on a substantial modification of the original method. This modification avoids the tedious, step-wise process of locating the threshold in successive trials, several of which are needed to determine the exact threshold intensity. This is accomplished by fixing the intensity of the stimulus and allowing it to act until a response results.

Thus, D'Amour and Smith,²¹ attracted by the report¹⁰ of the remarkable stability of the Hardy-Wolff pain threshold in man, set out to try the fixed intensity technique in rats. Their apparatus was the height of simplicity, in that it consisted of an auto headlight bulb focused by a reflector on a groove into which the tip of a rat's tail could be placed. Switching on the current started a stop-watch which was stopped when the rat flicked its tail away. Unfortunately, there is a limit on the amount of heat that may be applied without producing a burn. This necessitates cutting off the heat when analgesia is deep enough to prevent a response within a safe exposure time. The cut-off time is related, of course, to the

intensity of the stimulus, and most workers find that it is about $2\frac{1}{2}$ times the normal threshold time. One can be conservative in setting the cut-off time, since, with rats at least, there is little likelihood of obtaining a response at all if there has been none within twice the normal threshold reaction time. D'Amour and Smith cut off the stimulus at nine seconds. Others, using less intensity, have prolonged it up to 15 seconds.

With respect to the normal responses, the statement of D'Amour and Smith is worth quoting. They say, "We were greatly surprised at last to find a biologic reaction subject apparently to very little individual variation." Their rats, run twice a week, showed no evidence of conditioning. They give one protocol which permits calculation of the standard deviation of the normal threshold as being roughly ± 0.6 second or about 12 per cent of the average of 4.6 seconds. In our hands, the standard deviation is ± 0.27 second or about 8 per cent of our average of 3.5 seconds.

Interpretation of Data. With respect to handling the data of the D'Amour-Smith technique, it is interesting to note that, of the four groups of investigators who have used it, no two have interpreted their data in exactly the same way. D'Amour and Smith converted their results into the all-or-none type on the basis of the proportion showing "complete" analgesia within their cut-off time of about nine seconds. Ercoli and Lewis,²² in 1944, calculated what they termed the "average analgesic dose," which lends itself poorly to quantitative comparisons between drugs. Davies and his associates²³ found that the increase in reaction time in seconds plotted linearly against log dose and used this basis of effects. Foster and Carman²⁴ describe an "analgesic index" which is the square root of the ratio of the average maximum reaction time after the drug to the average pre-injection reaction time. They found that indices so obtained plotted against dose in approximately a straight line. Of these four methods, two measure the drug effects in terms of the increase in threshold, while the other two take the final level of reaction time as the better measure of effect. This lack of unanimity indicates fundamental differences in thinking.

All of the above authors encountered the problem peculiar to this method engendered by the frequent occurrence of such complete analgesia that the test animal fails to respond within the cut-off time. Thus, there results a set of data which is a hybrid mixture of graded and quantal responses. Thus far, this situation has been met by assuming that each animal actually reacted at the cut-off time and including these assumed values in the averages. This practice introduces a bias into the results which is less or great, depending on the actual magnitude of the arbitrary cut-off time. It is especially disturbing when the responses are figured in terms of the *increases* in reaction time above the normal. However, because of the simplicity of the D'Amour-Smith procedure as a whole, this problem is worthy of serious study.

To start with, there is the fundamental question of whether the reaction time of the medicated subject is correlated with that observed before dosing; that is, does a low normal threshold signify that the subject will also show a lower threshold after medication?

An answer to this question may be found in data obtained by Dr. Lewis of our laboratory,²⁵ using the D'Amour-Smith method, except that the stimulus was applied to the back instead of the tail. His experiment involved a comparison of methadone, levo-methadone, and a compound related to Demerol known to us as WIN 1539, and which chemically is 1-methyl-4-(3-hydroxyphenyl)-4-piperidyl ethyl ketone hydrochloride. The doses ranged from those producing just significant effects on the low side to those which made some, but not all, of the rats refractory on the high side. Four dosage levels were used which were spaced logarithmically in steps of 60 per cent increments. A group of 40 albino rats was assigned to each drug, and sub-groups of 10 rats received a different one of the four doses on each test day according to a balanced, latin square design. Only three readings were made on each rat each test day, namely, the normal threshold and at 30 and 60 minutes after dosing. To insure that each rat would be exposed to the same heat intensity, the filament voltage was adjusted to a constant heat output using a radiometer. The stimulus was cut off at 8 seconds.

Since Winder has shown that the frequency distribution of the normal thresholds of guinea pigs is more nearly normal on a logarithmic basis,¹⁵ our data were tested in this respect. While the distribution of the logarithms was slightly nearer normal, the difference was not great enough to justify the inconvenience of transforming the original data.

The responses of the individual rats to the drugs give the impression of marked heterogeneity. However, study shows that they are distributed in a fairly normal pattern with a few rats being sufficiently sensitive to the drugs that all three of the higher doses made them completely refractory to the pain stimulus, and with about an equal number showing relatively little analgesic effect from even the highest dose. There were some cases in which the effect from the lowest dose exceeded those of the three higher doses.

As a simple means of showing whether the observed effects were correlated with the respective normal thresholds, the latter were ranked in order from low to high. Then the respective post-injection reaction times of all average rats showing the same normal threshold were calculated. These averages represented the responses to all four doses of the same drug and are given in TABLE 3. It is clear that the reaction times after the drug are not correlated directly with the initial reaction time. The fourth column for each group shows the percentage increase in threshold due to the analgesic. These increases are greatest for the lower normal thresholds and indicate the bias that results from the apparently erroneous assumption that the *increase* in threshold is the true measure of analgesia.

TABLE 3
REACTION TIMES OF RATS AFTER MEDICATION, GROUPED ACCORDING TO THE RESPECTIVE NORMAL REACTION TIMES
(Normal reaction time—seconds)

Drug	3.0-3.1				3.2-3.3				3.4-3.5				3.6-3.7				3.8-3.9,*			
	No.		Sec. % Inc.		No.		Sec. % Inc.		No.		Sec. % Inc.		No.		Sec. % Inc.		No.		Sec. % Inc.	
	a	b			a	b			a	b			a	b			a	b		
Methadone	13	2	4.4	43	31	8	4.9	50	37	9	4.7	37	24	7	5.1	40	19	7	4.9	27
<i>l</i> -Methadone	14	6	4.7	53	18	9	4.7	43	43	7	4.6	34	34	17	5.3	45	8	4	4.9	28
WIN 1539	18	2	4.7	51	40	7	4.8	47	49	10	4.6	34	20	7	5.0	38	5	2	5.0	29

Number in column *a* includes only those rats which actually responded within the 8-second exposure time. Those in column *b* failed to react within the 8-second exposure time.

Values in third columns are average reaction times (rounded to nearest 0.1 second) at 30 minutes for all four doses of drug.

Values in fourth columns are percentage increases over the average reaction time for the group, i.e., over 3.05 sec. for the group showing normal times of 3.0 or 3.1 seconds.

* Only one rat showed a normal reaction time greater than 3.9 seconds.

Other data confirm this lack of correlation. This suggests that the final threshold may be a better measure of effect. Furthermore, considerable time might be saved in routine testing if one could omit the pre-injection normal threshold reading by insuring that the heat source is uniform and by following a cross-over experimental design. The practical question arises—can the initial reading be omitted? And further, what value accurately represents the effect? Is it the average, the median or some other parameter?

A careful study of the distribution of the individual reaction times following the three higher doses in the experiment just described, shows that there is such wide scatter that some doubt arises as to the propriety of striking even a simple average. This is particularly true for the highest doses, where about half of the animals proved refractory and failed to give finite reaction times.

Medians are sometimes useful in such cases and were determined, but it was found that they were not related to either dose or log dose in any obvious way. In fact, where more than half of the rats failed to react, as with the high doses of methadone and levo-methadone, the median is simply the eight-second cut-off time. The median can scarcely be used, therefore, as the measure of analgesic effect.

The simplest statistical device available for answering the questions just expressed is that of plotting the data to see which basis appears to give a family of parallel, straight lines best fitting the dosage-response data on all three compounds. Graphs of two different approaches are shown in FIGURES 3 and 4. The former shows the percentage increase of

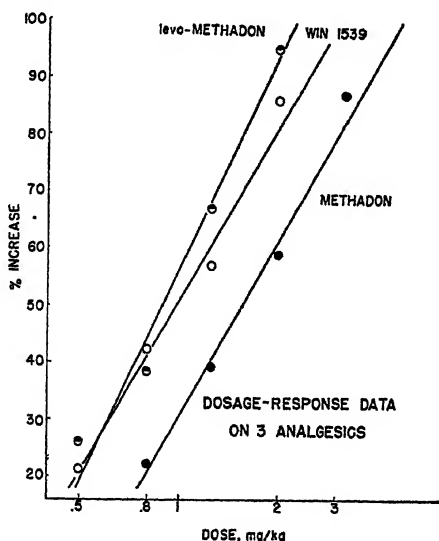


FIGURE 3.

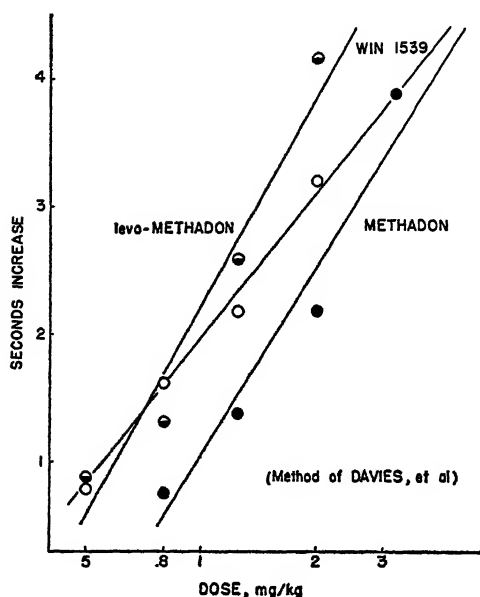


FIGURE 4.

the average threshold at 30 minutes after medication. While the points fit straight lines fairly well, it should be borne in mind that these data are biased in two ways. First of all, they are *increases*, and we have just seen that it is erroneous to assume that increases adequately reflect the analgesic effect. Second, the three highest points on each curve are partly assumed values since some rats were made completely refractory by the drug. FIGURE 4 shows the data on methadone and WIN 1539 plotted according to Davies²⁶ method, and here again the same objections hold.

Finally, we resorted to a slight modification of the treatment used originally by D'Amour and Smith. They took the proportion of rats which showed "complete" analgesia, that is, those absolutely refractory within the cut-off time. It seemed more logical to use the distribution of the normals as the basis for defining what may be termed "positive" analgesia. Thus, the tolerance limits^{26, 27} were calculated for the normal thresholds, and the point was picked beyond which less than 0.01 per cent of the normals would be expected to fall within a certainty of 95 per cent ($P=0.95$). The normals were sufficiently uniform for this dividing line to be the same for all the data, or about 1.4 seconds above the respective group average normal threshold. Thus, the values in TABLE 4 were obtained. These include the respective number of those considered "positive" and, of this group, those found to be "complete" or altogether refractory to stimulus within the 8-second exposure. The percentage "positive" is given in the next-to-last column.

TABLE 4

DOSAGE—EFFECT DATA ON METHADONE, LEVO-METHADONE AND WIN 1539;
D'AMOUR-SMITH METHOD ON RATS
(40 rats per dose)

Drug	Dose mg./kg.	Normal thresh. sec.	Analgesic effect			
			No. positive	No. complete	% Total	Probit
Methadone	.8*	3.49	4	0	10.3	3.74
	1.25	3.43	10	2	25.0	4.33
	2.0	3.44	27	6	67.5	5.45
	3.2*	3.47	38	25	97.4	6.94
levo-Methadone	.5	3.44	3	0	7.5	3.56
	.8	3.44	15	1	37.5	4.68
	1.25	3.51	33	9	82.5	5.93
	2.0	3.45	40	32	100	(7.50)
WIN 1539	.5	3.37	4	0	10.0	3.72
	.8	3.39	16	4	40.0	4.75
	1.25	3.39	26	8	65.0	5.39
	2.0	3.36	37	16	92.5	6.44

* Only 39 rats on these doses, due to the accidental death of one rat of the group.

Plotting the percentage of total "positives" against log dose gave curves with a strong sigmoidal trend, so that it seemed appropriate to rectify the curve by transforming the percentages to their corresponding probits. The resulting probit-log dose curves are shown in FIGURE 5. An arrow attached to the point for the highest dose on levo-methadone indicates a 100 per cent response for which the expected probit is shown, there being no finite probit for 100 per cent.

These curves involve no assumptions with respect to the initial threshold except that it is normally distributed, as may readily be shown. The cases of "complete" analgesia require no special consideration since they are surely "positive." However, although these curves may have a somewhat sounder theoretical basis, it must be admitted that, by visual inspection at least, the fit of the straight lines is not significantly better than that seen in FIGURES 3 and 4. Considering the reservations expressed regarding the validity of these curves, it seems scarcely justified to apply the conventional statistical tests for goodness-of-fit to them.

The graphical estimation of potency is the same for all of them, indicating that levo-methadone is about 70 per cent more potent than the racemic form and WIN 1539 is some 60 per cent stronger than *dl*-methadone. Using the recently described nomographic procedure of Litchfield and Wilcoxon,²⁸ the respective potencies of *l*-methadone and WIN 1539 are 177 and 158 per cent of that of racemic methadone. The confidence limits are about ± 30 per cent for $P=.95$, so that both drugs are clearly more potent than racemic methadone. The difference between *l*-methadone and WIN 1539 is not significant.

By stating these potency estimates, attention is focused on the fact

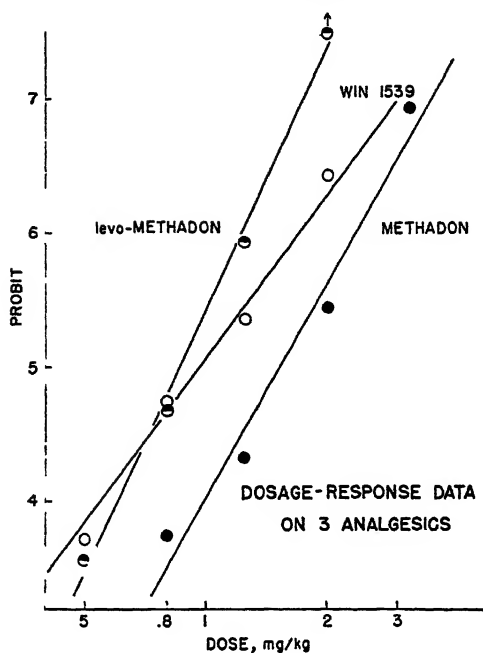


FIGURE 5.

that the ultimate object of an analgesic test is to obtain such an estimate along with some measure of its reliability, preferably the standard error. Since analgesia has two qualities, that of intensity as well as duration, it is conceivable that potency estimates might be based on either or both. Thus far, however, no one has worked out a means of combining measurements of both of these into a single parameter so that, if considered at all, they are taken separately.

While practically all of this discussion has been devoted to discussing the radiant thermal stimulus methods, there is another thermal stimulus method which deserves passing mention. Woolfe and Macdonald⁴ described the use of thermostatically heated zinc plates to provide a conducted thermal stimulus to the feet of mice. Dr. Eddy has kindly provided me with data obtained using this method on normal animals and on mice treated with Demerol and methadone.²⁹ Analysis of these data indicates that the mice vary greatly, not only from one to another but also between control readings taken 20 minutes apart. The odds that such discrepancies were due to normal chance variation are about one in a hundred. In view of this, and of the generally accepted fact that mice are about the most heterogeneous small laboratory animal known, present information does not permit rating the Woolfe-Macdonald method very high except as a convenient, rough screening method.

Summary

In summary, it appears that, of the three analgesic test methods which stand out, the tooth-pulp method has scarcely been studied enough to evaluate it fairly. The Hardy-Wolff-Goodell fixed-time method and its principal modification, the D'Amour-Smith fixed-intensity method, can be compared with the five criteria of the ideal method. Both rate well as to safety and universal applicability. The fixed-time method is the slower one when used on animals and therefore must be rated lower on the score of practicability. With human subjects, this is much less important. Both methods discriminate about equally, although any advantage is on the side of the D'Amour-Smith technique. Really good information on the element of reproducibility is lacking, except for the normal threshold determination where the fixed-time procedure is definitely superior. Neither method is sufficiently sensitive. The effects of mild analgesics such as aspirin are picked up only with difficulty and the maximum possible effect is limited.

From the standpoint of interpreting the data and making use of all the observations obtained, the complications are somewhat less formidable with the D'Amour-Smith method than with the original Hardy-Wolff-Goodell method. With both, there are fundamental problems still unsolved.

In the final balance, it would appear that, for comparing analgesics in laboratory animals, the D'Amour-Smith method is preferable; for human subjects, the original Hardy-Wolff-Goodell method is best.

Bibliography

1. GOETZL, F. R., D. Y. BURRILL, & A. C. IVY. 1946. A critical analysis of algometric methods with suggestions for a useful procedure. *Quart. Bull. Northw. Univ. Med. School* 17: 280-291.
2. HÄFFNER, F. 1929. Experimentelle Prüfung schmerzstillender Mittel. *Deutsch. med. Woch.* 55: 731-733.
3. SCHAUMANN, O. 1940. Über eine neue Klasse von Verbindungen mit spasmolytischer und zentral analgetischer Wirksamkeit unter besonderer Berücksichtigung des 1-Methyl-4-phenyl-piperidin-4-carbonsäure-äthylesters (Dolantin). *Arch. Exp. Path. & Pharmacol.* 196: 109-136.
4. WOOLFE, G., & A. D. MACDONALD. 1944. The evaluation of the analgesic action of pethidine hydrochloride (Demerol). *J. Pharm.* 80: 300-307.
5. BLISS, C. I., & E. L. SEVINGHAUS. 1947. A collaborative study of methods assaying analgesic drugs. *Fed. Proc.* 6: 310-311.
6. KOLL, W., & H. REFFERT. 1938. Eine neue Methode zur Messung analgetischer Wirkungen im Tierversuch—Versuche mit Morphin und einigen Morphinderivaten am Hund. *Arch. Exp. Pharm. Pathol.* 190: 687-711.
7. HARDY, J. D., H. G. WOLFF, & H. GOODELL. 1940. Studies on pain threshold: observations on spatial summation of pain. *J. Clin. Invest.* 19: 649-657.
8. WOLFF, H. G., J. D. HARDY, & H. GOODELL. 1940. Studies on pain. Measurement of the effect of morphine, codeine, and other opiates on the pain threshold and analysis of their relation to the pain experience. *J. Clin. Invest.* 29: 659-680.
9. HOLLAND, H. L., & E. G. GROSS. 1948. Two new analgesics. *Fed. Proc.* 7: 227-228.

10. SCHUMACHER, G. A., H. GOODELL, J. D. HARDY, & H. G. WOLFF. 1940. Uniformity of the pain threshold in man. *Science* 92: 110-112.
11. DENTON, J. E., O. H. STRAUS, W. E. WADDELL, & H. K. BEECHER. 1948. A comparison of side actions and analgesic effects of morphine, amidone and its isomers in man. *Fed. Proc.* 7: 215.
12. THORP, R. H. 1946. The assessment of analgesic activity in new synthetic drugs. *Brit. J. Pharm.* 1: 113-126.
13. DODDS, E. C., W. LAWSON, S. A. SIMPSON, & P. C. WILLIAMS. 1945. Testing diphenylethylamine compounds for analgesic action. *J. Physiol.* 104: 47-51.
14. WINDER, C. V., C. C. PFEIFFER, & G. L. MAISON. 1946. The nociceptive contraction of the cutaneous muscle of the guinea pig as elicited by radiant heat, with observations on the mode of action of morphine. *Arch. Int. Pharmacodyn.* 72: 329-359.
15. WINDER, C. V. 1947. Distribution of the resting pain-reaction thresholds in guinea-pigs, with a statistical concept of gradation of biological effect. *Yale J. Biol. Med.* 19: 289-310.
16. WINDER, C. V. 1947. A preliminary test for analgetic action in guinea pigs. *Arch. Int. Pharmacodyn.* 74: 176-192.
17. WINDER, C. V. 1947. Quantitative evaluation of analgetic action in guinea pigs. *Arch. Int. Pharmacodyn.* 74: 219-232.
18. BONNYCASTLE, D. D., & J. MOLLAND. 1948. An examination of potentiating action of aspirin upon codeine in raising the pain threshold. *Fed. Proc.* 7: 206.
19. MATHER, K. 1946. *Statistical Analysis in Biology*. Interscience Publishers, Inc. New York.
20. WOLFF, H. G. 1945. Cornell Conference on Therapy: Psychologic aspects of treatment of pain. *N. Y. State J. Med.* 45: 1003-1009.
21. D'AMOUR, F. E., & D. L. SMITH. 1941. A method for determining loss of pain sensation. *J. Pharm. & Exp. Therap.* 72: 74-79.
22. ERCOLI, N., & M. N. LEWIS. 1945. Studies on analgesics. I. The time-action curves of morphine, codeine, dilaudid and demerol by various methods of administration. II. Analgesic activity of acetyl salicylic acid and aminopyrine. *J. Pharm. & Exp. Therap.* 84: 301-317.
23. DAVIES, O. L., J. RAVENTÓS, & A. E. WALPOLE. 1946. A method for the evaluation of analgesic activity using rats. *Brit. J. Pharm.* 1: 255-264.
24. FOSTER, R. H. K., & A. CARMAN. 1947. Studies in analgesia: Piperidine derivatives with morphine-like activity. *J. Pharm. & Exp. Therap.* 91: 195-209.
25. LEWIS, J. R. Unpublished data.
26. WALD, A., & J. WOLFOWITZ. 1946. Tolerance limits for a normal distribution. *Ann. Math. Stat.* 17: 208-215.
27. EISENHART, C., M. W. HASTAY, & W. A. WALLIS, Editors. 1947. *Techniques of Statistical Analysis*. McGraw-Hill. New York.
28. LITCHFIELD, J. T., & F. WILCOXON. 1948. A simplified method of evaluating dose-effect experiments. *Fed. Proc.* 7: 240.
29. EDDY, N. B. Personal communication.

PHARMACOLOGY OF METOPON AND OTHER NEW ANALGESIC OPIUM DERIVATIVES

By NATHAN B. EDDY

National Institute of Health, Bethesda, Maryland

MORPHINE has always been, and continues to be, the standard for comparison in the field of analgesia in spite of its many disadvantages—irregularity of action of oral doses, dopiness (as a rule) with adequate pain relief, high incidence of side reactions, particularly nausea and vomiting, narrow margin between analgesic and respiratory depressant doses, and the property of rapid development of tolerance and addiction. The most intensive attack upon these disadvantages began about 20 years ago in the coordinated program of the Drug Addiction Committee of the National Research Council. It advanced under those auspices through 1939 and is still going forward at the National Institute of Health. Hundreds of morphine derivatives have been made and examined pharmacologically, to try to find the group, or groups, in the morphine molecule associated with its various actions and the modification of those groups which would result in the most satisfactory clinical effectiveness. The data collected permit a number of generalizations:

- (1) Modification (muzzling) of the phenolic hydroxyl decreases and substitution or removal of the alcoholic hydroxyl increases morphine-like activity.
- (2) Removal, by hydrogenation, of the double bond adjacent to the alcoholic hydroxyl usually increases activity.
- (3) Opening the nitrogen ring practically abolishes morphine-like activity.
- (4) Attachment of an alkyl group to the hydroaromatic ring results in quantitative dissociation of morphine-like effects, increasing some and decreasing others.
- (5) No active morphine derivative has been made which is entirely free of addiction liability.

A description of a few of the newer morphine derivatives will illustrate, in some measure, these generalizations, and will permit an appraisal of the extent to which progress has been made toward the goal of greater clinical effectiveness.

About 25 years ago, German chemists made the ketone derivatives of morphine and codeine, namely Dilaudid and Dicodid. The former was introduced in this country in 1932; the latter quite recently, both as Dicodid and Hycodan. FIGURE 1 shows the structure of these compounds and their relation to morphine and codeine, respectively, and TABLE 1 gives the minimal doses for some of the more important effects. With both the morphine and the codeine, analog toxicity and activity have been increased by hydrogenation and by the introduction of the ketone, and to about the same extent. In other words, although the ef-

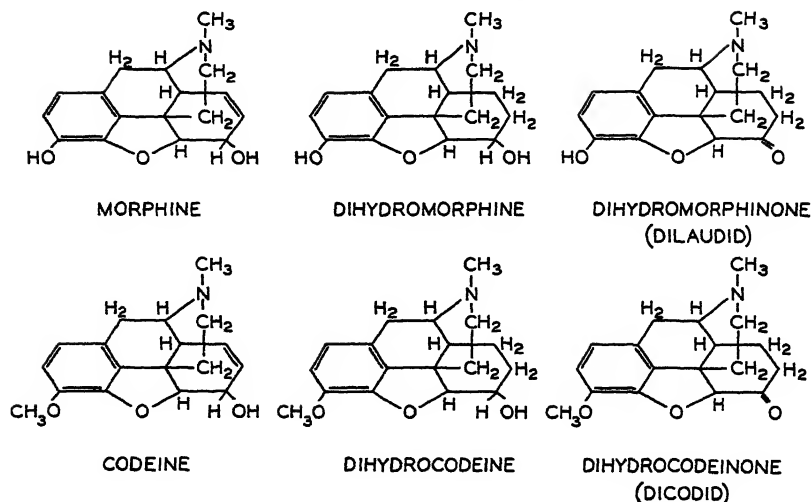


FIGURE 1.

fective analgesic dose has been greatly reduced in Dilaudid and Dicodid, the dose for the production of other less desirable effects has been reduced similarly. Also, the properties of tolerance, of dependence production, of pain relief, and of anti-tussive action have changed in the same direction. Dilaudid, then, offers little advantage over morphine in any respect, except reduction in the size of the dose to produce an effect. Dicodid has been used in this country principally for cough relief, and a number of reports have appeared comparing it with codeine for that

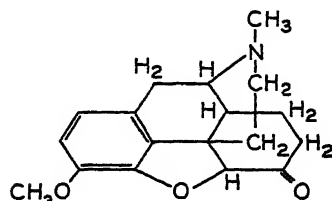
TABLE I
EFFECTIVE DOSES IN ANIMALS

	<i>LD₅₀</i> <i>Mouse</i>	<i>Depressant</i> <i>Rat</i>	<i>Analgesic</i> <i>Cat</i>	<i>Emetic</i> <i>Cat</i>	<i>Respiratory</i> <i>Rabbit</i>	<i>Intestinal</i> <i>Rabbit</i>
Morphine	581	4.50	0.75	0.22	0.15	4.54
Dihydromorphine	133	7.90	0.26	0.17	0.11	3.54
Dihydromorphinone	84	0.80	0.17	0.08	0.01	0.62
Codeine	241	16.00	8.04	16.00	1.30	16.00
Dihydrocodeine	225	14.10	7.20	3.22	0.90	5.80
Dihydrocodeinone	86	4.20	1.28	2.56	0.08	3.40

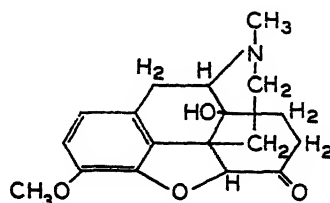
purpose. These reports have denied the development of tolerance and dependence and have claimed good anti-tussive effect with Dicodid. Cases of addiction to Dicodid were described in Germany where the drug was more widely used, much more frequently, in fact, than one would expect with similar administration of codeine. The risk of addiction, even with morphine, is somewhat reduced by keeping the dose at the minimal level for symptomatic relief, so that the small doses of Dicodid which are used for cough may be attended by less risk of addiction than the amounts which were used by the Germans for other purposes. All

the evidence indicates, however, that the risk is materially greater than with codeine, and recognition of this fact is essential, even though the use of Dicodid is limited to the treatment of cough.

The Germans made another ketone derivative of codeine, namely, dihydrohydroxycodeinone, which they named Eucodal (FIGURE 2). It is



DIHYDROCODEINONE
(DICODID)



DIHYDROHYDROXYCODEINONE
(EUCODAL)

FIGURE 2.

most closely related to Dicodid, differing from the latter by the introduction of a hydroxyl group at C-14. Experimentally, Eucodal is less toxic than Dicodid, but it is more emetic and not materially different in other respects. Reports in the German literature, and our own experience in its substitution for morphine in known addicts, indicate that it has a high degree of addiction liability, probably comparable to that of morphine itself. Eucodal has not been introduced in this country, and it is doubtful that anything will be gained by its employment.

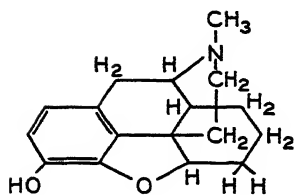
By complete removal of the alcoholic hydroxyl of morphine and its substitution by hydrogen (FIGURE 3), analgesic effect has been increased as much, or more, than in any other morphine derivative (TABLE 2). This

TABLE 2
EFFECTIVE DOSES IN ANIMALS

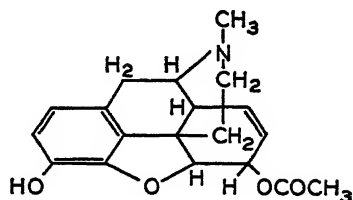
	<i>LD₅₀</i> <i>Mouse</i>	<i>Depressant</i> <i>Rat</i>	<i>Analgesic</i> <i>Cat</i>	<i>Emetic</i> <i>Cat</i>	<i>Respiratory</i> <i>Rabbit</i>	<i>Intestinal</i> <i>Rabbit</i>
Morphine	531	6.75	0.75	0.23	0.15	4.54
6-Acetylmorphine	293	1.8	0.18	0.18	0.019	0.8
Dihydromorphine	183	7.9	0.26	0.17	0.11	3.54
Dihydrodesoxymorphine-D	104	0.32	0.08	1.0+	0.012	0.32
6-Acetyldihydromorphine	99	6.3	1.35	0.90	0.072	1.8

has been established clinically as well as experimentally. The emetic effect of Desomorphine is slight, but the duration of its analgesic action is short, not more than half that of morphine, its respiratory depressant effect is exaggerated as much as its analgesic action, and its addiction liability is, perhaps, as great as that of morphine.

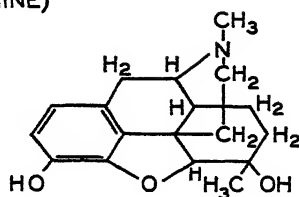
During the last war, when it seemed possible that our morphine supply might become short, attention was focused for a time on 6-acetyl-



DIHYDRODESOXYMORPHINE-D
(DESMORPHINE)



6-ACETYLMORPHINE



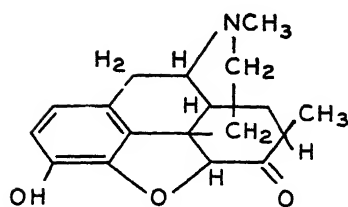
6-METHYLDIHYDROMORPHINE

FIGURE 3.

morphine (FIGURE 3). It can be produced in quantitative yield from morphine by very simple means, and the product is 4 times more effective as an analgesic agent. Unfortunately, all morphine-like effects are increased to about the same extent, so that the only gain resulting from monoacetylation of morphine seems to be that one obtains 4 analgesic doses for each one entering into the reaction. 6-Acetyldihydromorphine (TABLE 2) is an exception to the generalization that hydrogenation of 7-8 double bond increases morphine-like activity. The compound is less effective in sedative, analgesic, and respiratory action than 6-acetylmorphine.

6-Methyldihydromorphine is a new type of modification of the morphine molecule. The alcoholic hydroxyl is retained and a methyl group is substituted for hydrogen on the same nuclear carbon. Analgesic effect is exhibited by this compound to the same extent as by morphine itself, and the duration of this effect has been increased almost twice. Most noteworthy is the fact that 6-methyldihydromorphine seems to have almost no effect on the withdrawal symptoms of known morphine addicts. The first and simplest test, in the assay of addiction liability in any new agent, is administration of a single dose at the peak of the morphine abstinence syndrome, and resulting suppression of the syndrome is considered evidence of addiction-sustaining potency. Absence of such suppression with 6-methyldihydromorphine indicates at least that it has low addiction-sustaining power and suggests that it may have low addiction-producing liability.

Methyldihydromorphinone or Metopon (FIGURE 4) is another new type of morphine derivative in which, in addition to hydrogenation and

METOPON

(methyl dihydromorphinone)

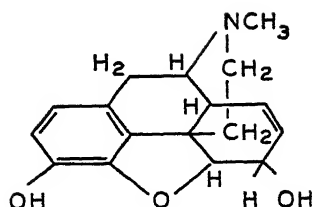
MORPHINE

FIGURE 4.

substitution of the alcoholic hydroxyl by oxygen, a new substituent, a methyl group, has been attached to the hydroaromatic ring of the morphine molecule. The last is the essential feature, because in comparison with dihydromorphinone (TABLE 3) Metopon exhibits an exaggeration

TABLE 3
EFFECTIVE DOSES IN ANIMALS

	<i>LD₅₀</i> <i>Mouse</i>	<i>Depressant</i> <i>Rat</i>	<i>Analgesic</i> <i>Cat</i>	<i>Emetic</i> <i>Cat</i>	<i>Respiratory</i> <i>Rabbit</i>	<i>Intestinal</i> <i>Rabbit</i>
Dihydromorphinone	84	1.77	0.17	0.08	0.011	0.62
Methyl dihydromorphinone	25	3.00	0.07	0.15	0.012	3.00

of analgesic effectiveness and diminution of sedative, euphoric, emetic, and intestinal actions. Animal experiments and preliminary clinical trial established these differences and, furthermore, indicated that oral and parenteral analgesic doses were essentially the same. With prolonged administration, tolerance and dependence developed more slowly with Metopon than with morphine. These results suggest that Metopon is an analgesic agent particularly advantageous for oral administration for chronic pain, since its action should be attainable with relative freedom from side effects and without the disadvantages of hypodermic injection and of loss of effectiveness by rapid development of tolerance.

About two years ago, the Committee on Drug Addiction of the National Research Council recommended limited production of Metopon and arranged for its distribution for one year under its supervision. The Committee intended a large-scale clinical experiment in which the drug would be made available to one class of patients, cases of terminal malignancy, and reports would be gathered on its usefulness. Acting for the Committee, the author has screened all Metopon orders and has analyzed and tabulated the physicians' reports. The drug has gone to more than five thousand cancer patients, and reports have been received on its trial in more than half of that number.

TABLE 4 summarizes the results as shown on all reports, disregarding dosage and other disturbing factors. Pain relief has been fair or better

TABLE 4

Pain relief with Metopon, all reports	
Complete	32.3%
Fair	41.8
Poor	18.0
None	7.8
Mental condition after Metopon	
Clear	74.2%
Dull	18.8
Dopey	5.2
Confused	1.6
Sleep promoted, or permitted, by relief of pain	27.0

in 74 per cent of all trials, and mental clarity without tendency to sleep has been reported in an equal number. Side reactions (TABLE 5) have been entirely absent in 90 per cent. Nausea, the most frequent and most

TABLE 5

SIDE REACTIONS WITH METOPON

None	2096
Nausea	142
Restlessness	28
Headache	17
Perspiration	18
Dryness or burning in throat or stomach	21
Dizziness	13
Euphoria	10
Disturbing dreams	9
Burning of skin and/or hot flashes	7
Itching	6
Anorexia	5
Urticaria	2
Tinnitus	3
Allergic reaction	2
Difficulty in urination	2
Palpitation	1
Hiccough	1
General numbness	1
Nausea, previously reported (376), persisting after Metopon	36
Signs of physical and psychic dependence for which Metopon was admittedly discontinued after brief trial	138
Signs of physical dependence disappearing under continued Metopon administration	52

disturbing side reaction, has been reported 142 times, but was evident in only 36 out of 376 patients for whom its occurrence after other medication is known to have been a reason for trying Metopon. Other side reactions have been minor and infrequent.

The greatest difficulty encountered in the trial of Metopon stems from the amount of previous medication for pain relief. Among the cases reported, Metopon has been tried in little more than 100 who had re-

ceived no previous narcotic; in about 500 who had received narcotics for 6 weeks or less, a period in which some but relatively little tolerance and dependence will have developed as a rule; and in four times that number who had received narcotics for 2 months or more, when tolerance and dependence almost always will have developed. Using only the results of doses of 6 mg., the recommended minimal analgesic dose, these have been grouped to determine the influence of the previous medication (TABLE 6). Quite definitely, the more previous narcotic medication the

TABLE 6
INFLUENCE OF PREVIOUS NARCOTIC ADMINISTRATION
ON RESULT OBTAINED WITH METOPON
(6 mg. per dose)

	<i>No previous narcotic</i>	<i>Narcotic administration 6 weeks or less</i>	<i>Narcotic administration 2 months or more</i>
Number of cases	48	213	665
Pain relief:			
Complete	27	93	176
Fair	17	57	244
Poor	1	40	160
None	3	23	85
	91 %	70 %	63 %
Incidence of nausea	0	9	30

patient has received, the less will be the incidence of adequate relief. When tolerance to morphine and related substances has developed, some degree of cross-tolerance to Metopon may be present, diminishing somewhat its effectiveness.

It is significant that nausea after Metopon has not been reported in the patients who had received no other narcotic. This would indicate that much of the nausea reported among the side reactions was related to the previous narcotic administration, a withdrawal sign, and was not a direct effect of Metopon. The drug not only is slower in producing dependence but also incompletely supports an established dependence. The switch from morphine, etc., to oral Metopon can be and has been made satisfactorily in many of the cases reported, many times in spite of the temporary appearance of mild abstinence phenomena. It will be effected most easily, probably, by rapid reduction of the other drug during the first few days of Metopon administration.

More than 300 physicians have submitted two or more reports on the same patient, covering periods of administration of 3 to 20 weeks. At least two-thirds of these cases have shown no indication of development of tolerance to Metopon.

Metopon is difficult and expensive to make and will probably continue to be in limited supply. Consequently, it should not be considered a substitute for, or competitor with morphine for hypodermic administrations for pre- and postoperative cases, for example. Rather, its ad-

vantages adapt it particularly to the ambulatory or semi-ambulatory patient with chronic pain. Indeed, in this field it stands out among morphine derivatives and other current agents of like potency as an analgesic for oral administration. Its use should be restricted, of course, to those cases in whom pain is of such severity as to demand an analgesic of morphine-like effectiveness. Especially if its use is started before tolerance to other narcotics has developed, it can reasonably be expected to give adequate pain relief for a longer period, not in hours per dose, but in weeks of comfort without notable tolerance and dependence, and to give that relief almost always without side reactions or dopiness.

PHARMACOLOGY OF DEMEROL AND ITS ANALOGUES

By FREDRICK F. YONKMAN*

Research Department, Ciba Pharmaceutical Products, Inc., Summit, N. J.

IN presenting the pharmacology of Demerol we have been forced, by limitations of space, to restrict the extent of our discussion, and we have selected from the numerous reports concerned with the pharmacology of this compound those features which we deemed to be the most important.

At least three of the previous papers have cited the fact that Demerol was the result of a search for a new synthetic atropine substitute. When Eisleb and Schaumann¹ synthesized Demerol back in 1939, they were not in search of a new analgesic agent but were seeking, rather, a spasmolytic agent which would act anticholinergically in the manner typical of atropine. The synthetic chemist reminds one of the hopeful fisherman. Frequently he goes out with the best equipment but all too often returns with very small fish. Then again, another may launch forth on his fishing expedition with a meager amount of equipment and facilities and come back with the prize, just as Eisleb and Schaumann did. Too often most of us go out full of hope and expectancy only to experience the pulling up of redundant material such as a lead boot or useless items comparable to a discarded tire; but apparently that is what keeps us going—the unexpected—and in time, we keep telling ourselves, some of us may be able to duplicate the experience of Eisleb and Schaumann.

There is some similarity, from a structural point of view, between Demerol, atropine, and morphine. A glance at the morphine structure (FIGURE 1) indicates that it is not a phenanthrene derivative as had long been supposed, but, according to Gulland and Robinson,² it is a derivative of the piperidine nucleus. In this sense, Demerol is akin to morphine since it, too, is a piperidine derivative; actually, by name, it is 1-methyl-4-phenylpiperidine-4-carboxylic acid. The similarity in terms of piperidine derivatives probably could account for the analgesic action of both Demerol and morphine but certainly could not be held accountable for the smooth-muscle actions of the respective compounds, since they are distinctly at variance in that morphine is spasmogenic whereas Demerol, as a rule, is spasmolytic. This point will be elaborated upon later. Reference to the atropine structure in FIGURE 1, if one is permitted to follow the pencil gymnastics of Schaumann, discloses that atropine could also be considered a close chemical relative of Demerol and, therefore, one could conclude that the somewhat similar pharmacological actions in terms of spasmolysis and anticholinergic activity might have been an-

* Lecturer in Pharmacology and Therapeutics, College of Physicians and Surgeons, Columbia University, New York, N. Y.

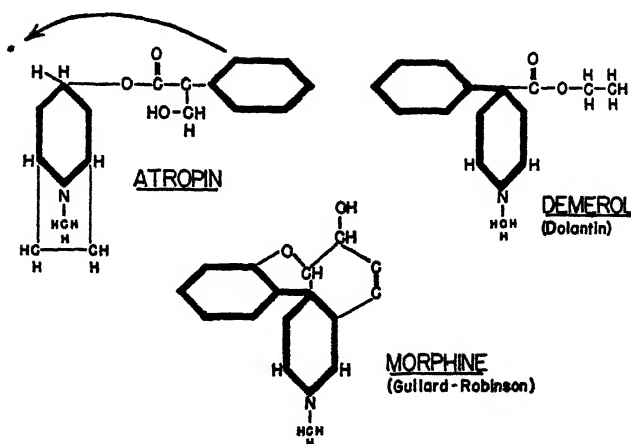


FIGURE 1. Atropine, Demerol (Dolantin), and morphine (SCHAUMANN).

ticipated. However, it is difficult to correlate the different types of central nervous system activity of Demerol and atropine on the basis of their chemical structures. Again, which feature of a chemical formula determines this type of activity or that, is indeed a fact difficult to prognosticate. We see certain resemblances and then hope for some degree of correlation. The latter is permissible up to a point; but extensive speculation is merely that, and frequently it leads ultimately to confusion.

Considerable pharmacological investigation of Demerol was made in this country soon after its introduction on the European continent. Along with the work of Climenko³ and Barlow⁴ in the Winthrop Laboratories at Rensselaer, we have the very important work emanating from university laboratories under the name of Gruber and his associates, Hart and Gruber, Jr.⁵ TABLE I indicates the relative safety of Demerol in three species of experimental animals, namely, white mice, albino rats, and rabbits:

TABLE I

<i>Animal</i>	<i>Route of administration</i>	<i>LD₅₀ mg./kg.</i>
White mice	{ Intraperitoneally	147
	{ Orally	221
Albino rats	Intraperitoneally	93
Rabbits (average weight 1.7 kg.)	Intravenously	32
Rabbits (average weight 3.1 kg.)	Intravenously	20

This work substantially agrees with that of several other investigators, and clinical experience with Demerol during the last several years supports the contention made earlier by various investigators that Demerol should prove to be a safe drug from the point of view of toxicity.

Gruber and his associates were eager to determine the acute effects of

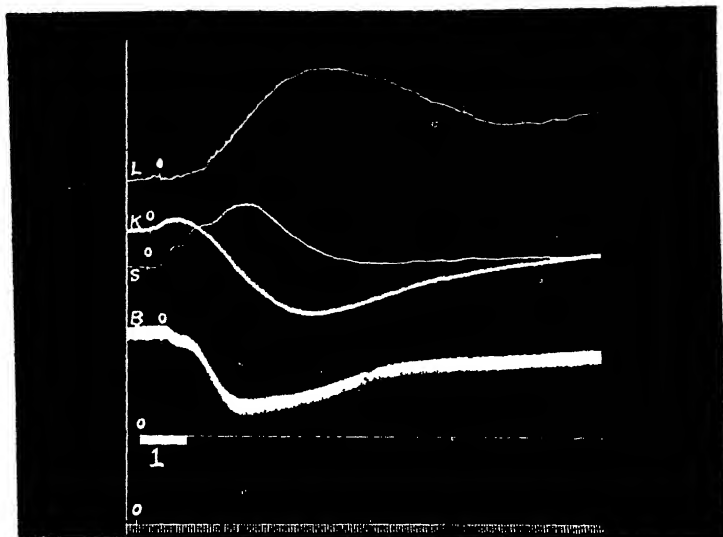


FIGURE 2. Dog 13 kg. ♀. Ether anesthesia. Bottom record, time in intervals of 6 seconds and zero blood pressure; above it, time and speed of injection of the drug, 50 mm. mercury pressure. *L*, volume record of rear extremity; *K*, kidney volume; *S*, volume of spleen; *B*, arterial blood pressure with mercury manometer. 0 indicates simultaneous points of all the levers. Up movement in *L*, *K*, and *S* records indicate increases and down movement decreases in volumes of the organs. At 1, 5.0 mg. of Demerol per kilogram were injected intravenously.

Demerol when administered intravenously to the anesthetized animal such as the dog. There resulted (FIGURE 2) a marked hypotension, of not too long duration, which was associated with an increase in leg volume and spleen volume, apparently at the expense of kidney volume, which was decreased. In other words, there was a redistribution of the blood associated with the acute hypotension, following the administration of Demerol. These data early brought the attention of the enterprising clinician to the fact that intravenous administration of Demerol might be associated, in the clinical patient, with marked symptoms of toxicity related to the acute hypotension. This will be referred to later when we discuss the clinical pharmacology of Demerol.

Demerol has recently been compared with methadone with reference to their effects on the blood pressure and respiration of the anesthetized dog. Demerol and methadone (FIGURE 3) are both depressing to blood pressure and likewise depressing to respiration. But the experiment cited, one which has been frequently and amply confirmed by others, demonstrates the propensity of methadone to depress respiration. This variance is a fact which becomes of considerable importance clinically, especially in the field of obstetrics, which will be referred to later.

During our early experiences⁶ with Demerol while at Wayne University in Detroit, we were impressed with the fact that this drug had exhibited important antihistaminic properties. This was detected in vari-

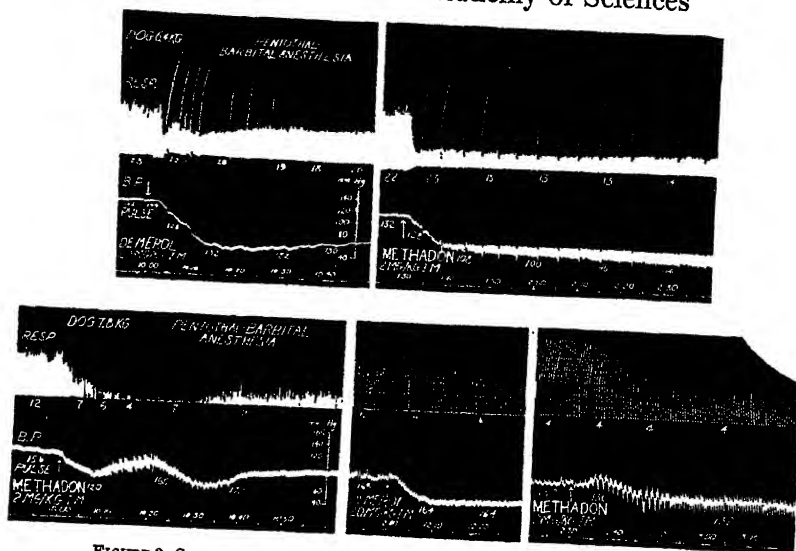


FIGURE 3. Comparison of the respiratory and circulatory effects of Demerol and methadone in anesthetized dogs upon intramuscular injection. (From TAINTER.¹⁵)

ous types of experiments in which the effects of several drugs on the nictitating membrane, salivation, pupil, and blood pressure were obtained in cats under Dial-urethane. The anticholinergic as well as the antihistaminic properties of Demerol and atropine have been demonstrated in more recent experiments (FIGURE 4) with the cooperation of Mr. Frank Roth. The secretagogue or sialogogic effect of 10γ of histamine was reduced from 4 cm. on the scale to .85 cm. after 500γ of Demerol had been administered intravenously. This amount of Demerol had no appreciable effect on the chorda tympanica control of salivation, whereas only 10γ of atropine completely nullified the secretory stimulation of 10γ of histamine and also completely eliminated the effects of faradizing the chorda tympanica nerve. In other words, Demerol was partially anticholinergic but definitely antihistaminic, whereas atropine exhibited strong potentialities in terms of both functions, that is, the humoral and the neural control of salivation.

Spasmolysis

The antihistaminic effect of Demerol is further demonstrated in FIGURE 5. The spasmolytic effect of Demerol against the spasmogenic effect of pitocin and epinephrine is likewise illustrated. Oddly enough, Demerol, in certain situations, seems to exert a spasmogenic action as is seen in the first tracing of FIGURE 5. Apparently, the effect to be observed following the addition of Demerol depends to some extent upon the functional state of the organism being tested. This finding has been amply

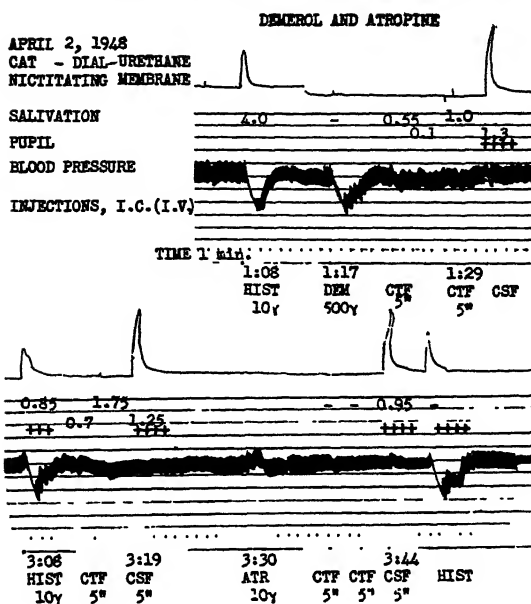


FIGURE 4. Demerol and atropine.

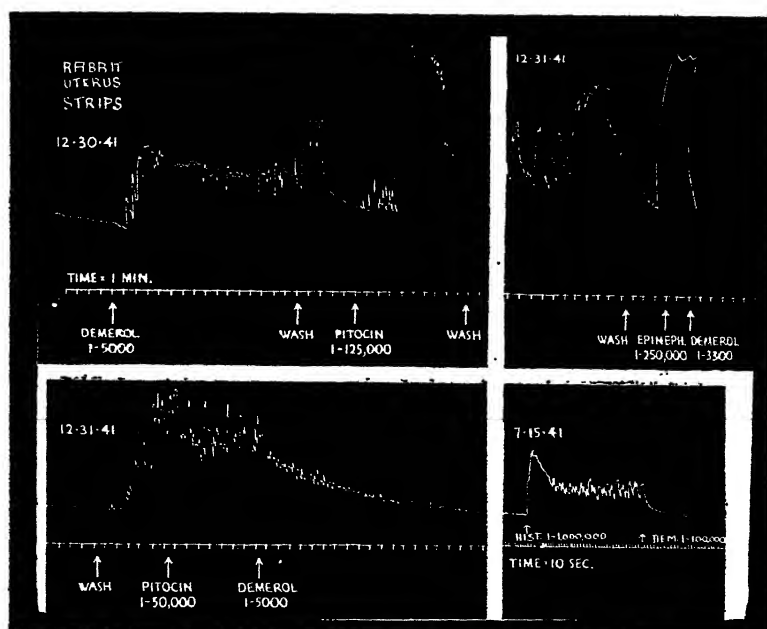


FIGURE 5. Rabbit uterus strips.

observed by others and we wish to report here complete corroboration of this fact as first reported by Gruber and his associates.

Because of the important contributions Dr. Gruber had made when he corroborated the work of my preceptor in pharmacology, Dr. O. H. Plant,⁷ in reference to the spasmogenic action of morphine on the intestine, it was highly desirable that we determine the character of the action of Demerol in this regard. FIGURE 6 illustrates the spasmogenic

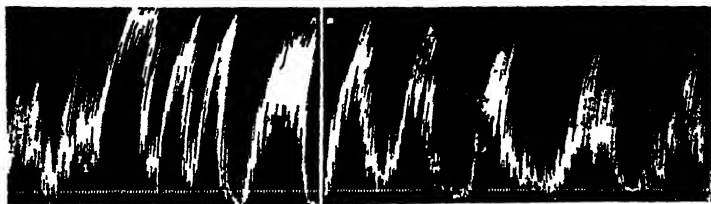


FIGURE 6. Reduced to $\frac{1}{2}$ original size. Dog 10.4 kg. ♀. Unanesthetized with Thiry-Vella fistula of ileum. Top record is of intestine and bottom record is the time in intervals of 15 seconds. Upstroke indicates contraction of muscle and increased tonus. At 1, 2.0 mg. per kilogram of Demerol were injected intravenously. For publication purposes, 15 minutes of the control record have been omitted and 5 minutes between records A and B.

type of action manifested by Demerol when injected intravenously in the dog prepared with a Thiry-Vella loop whose activity was registered kymographically by the so-called closed-balloon manometric method. Gruber's work in this regard has been amply confirmed by the work of many investigators including that of our associates.⁶ Because of the variance which frequently prevails between the results obtained experimentally in animals and in man, we were most eager to determine the effect of Demerol on the gastrointestinal tract of man. FIGURE 7 demonstrates the arrestive or quieting action of Demerol (S-140) on the intestinal tract of a patient prepared with an ileostomy. Peristaltic activity

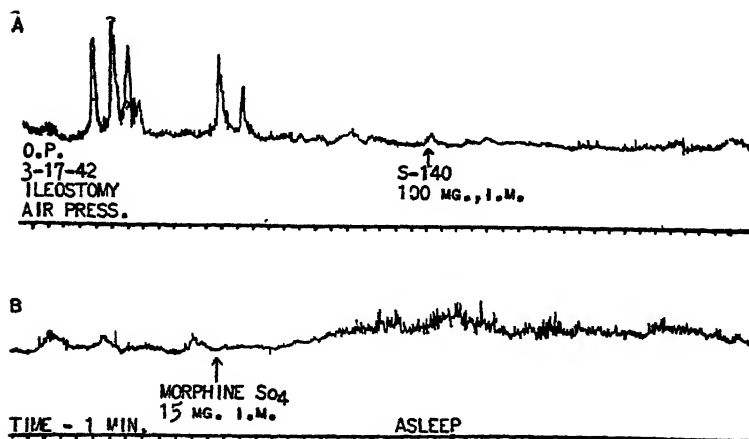


FIGURE 7. Demerol in man.

(or the large tonal waves) as seen in the control period was either prevented or decidedly moderated by Demerol, whereas morphine was spasmogenic in its typical manner. Interestingly enough, Demerol was not successful in preventing the spasmogenic activity of morphine in this patient, a finding different from that which frequently prevailed in the colon. FIGURE 8 presents a comparative study of morphine and Demerol

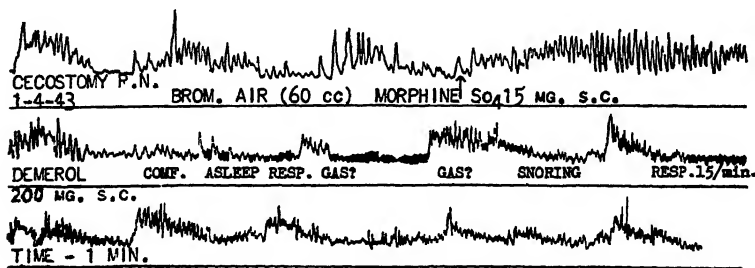


FIGURE 8. Demerol and morphine in man.

on the colon made in a patient afflicted surgically with a cecostomy. One notes that morphine was spasmogenic in this patient, whereas Demerol was spasmolytic in spite of the colonic spasticity prevailing after the opiate.

We are indebted to Dr. Batterman⁸ for his excellent studies on the effect of Demerol on the stomach. Dr. Batterman has made one of the finest investigations of Demerol, both experimentally and clinically, in this country, and it is a pleasure to pay him this tribute. One notes (FIGURE 9) that Demerol invariably produced a quiescent effect on the normal contractile waves prevailing in the stomach of man. This effect prevailed for varying periods of time, depending upon the patient and the dosage

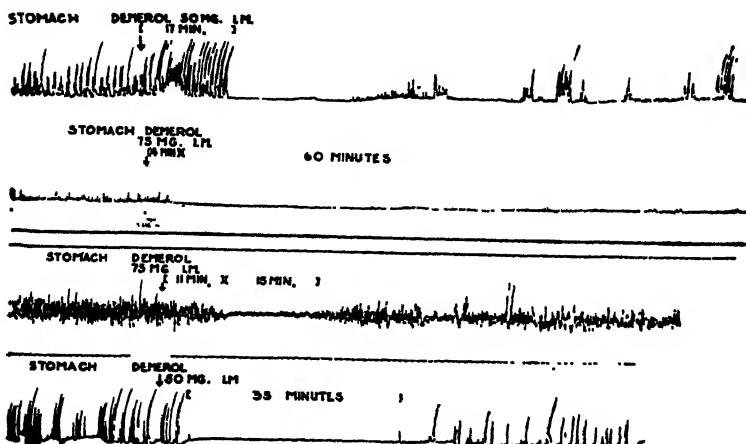


FIGURE 9. Demerol. (From BATTERMAN.⁸)

employed, but it is important to note that in no instance did Demerol produce a spasmogenic effect similar to that which it produces in the experimental dog or like that which morphine produces in man. This is a fact of great importance from a clinical point of view.

There remained one portion of the alimentary tract to be studied in detail, namely, the duodenum and the jejunum, and for these studies we are indebted to Dr. Gaensler and his associates¹ at the Boston City Hospital, who employed the technique designed by McGowan, Butsch, and others¹ at the Mayo Clinic. FIGURE 10 serves to illustrate the technique

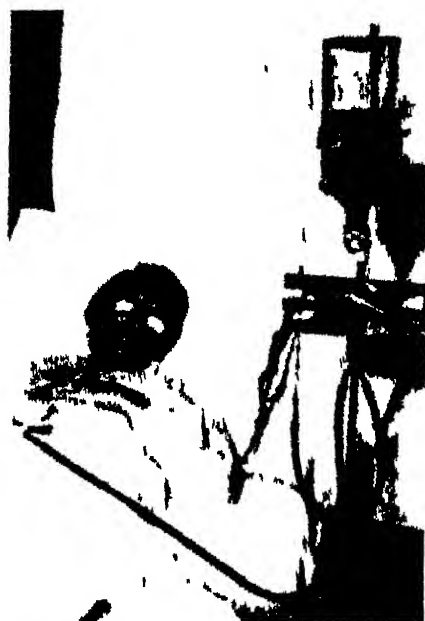


FIGURE 10. Apparatus connected to T-tube. This picture was taken shortly after administration of morphine; the fluid in the manometer has started to rise.

employed. Following cholecystectomy or removal of the gall bladder in certain patients, a T-tube was left in the common bile duct for purposes of drainage. At appropriate periods, this tube was connected manometrically to a saline or water flow system, the elevations and falls of which could be registered kymographically with a tambour system. The effects of various drugs are readily observable in FIGURES 11 and 12. One notes that amyl nitrite, following inhalation, relaxes the tone of the duodenum or jejunum whereas Demerol is markedly spasmogenic, much as it is in the ileum or small intestine of the dog.

Added evidence concerning the spasmogenic properties of Demerol is gained from cholangiograms of the patients studied with the use of diodrast instilled directly into the common duct. Following Demerol, no

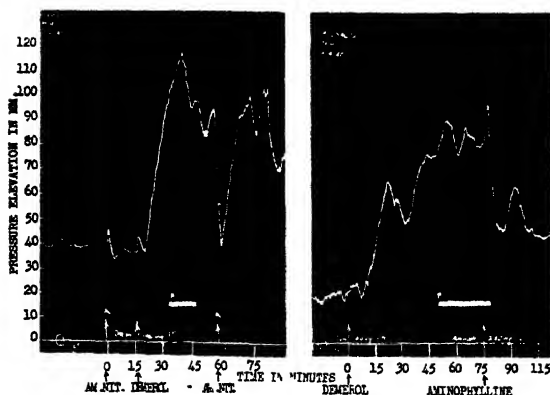


FIGURE 11. Kymograph tracings showing pressure changes associated with two instances of biliary colic produced in different patients by the administration of 100 mg. of Demerol intramuscularly. Relief in the first case was obtained with inhalations of amyl nitrite and in the second with 240 mg. of theophylline with ethylenediamine (aminophylline) given intravenously. *P*, pain. The base line or zero level in this and succeeding illustrations is at the level of the xiphoid with the patient in the prone position. (From GAENSLER *et al.*)

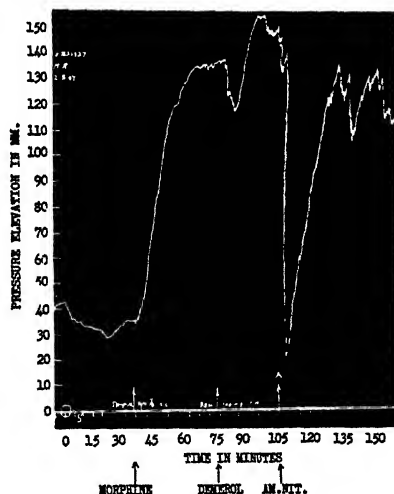


FIGURE 12. Demonstration of the inability of Demerol to relieve morphine-induced spasm. In this particular patient, Demerol produced more spasm than did morphine. Relief was obtained with amyl nitrite. (From GAENSLER *et al.*)

diodrast passed through the sphincter of Oddi into the duodenum (FIGURE 13), whereas inhalations of amyl nitrite relaxed this spasticity to such an extent that diodrast, as a contrast medium, did pass through the sphincter appreciably to reduce the amount contained in the common duct.

When one reviews the compilation of studies of ten patients with



FIGURE 13. A, cholangiograms: there is marked spasm at the lower end of the common bile duct with dilatation above hydrohepatosis. No diodrast reached the duodenum. B, the second plate was taken one minute following inhalation of amyl nitrite. The sphincter has relaxed and contrast media is flowing freely in to the duodenum. From GAENSLER *et al.*

T-tube drainage of the common bile duct under the influence of codeine, Demerol, and morphine (FIGURE 14), one is impressed with the close parallelism between the spasmogenic action of Demerol and morphine. In almost every case, codeine was not as spasmogenic in its action as were

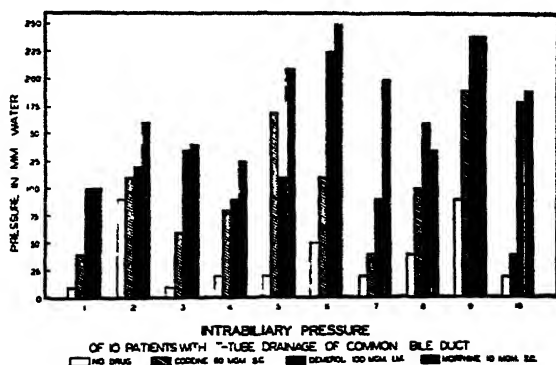


FIGURE 14. Comparison of the spasmogenic action of codeine, Demerol, and morphine on the sphincter mechanism of the common bile duct as measured by changes in the intrabiliary pressure in man. (From GAENSLER *et al.*)

Demerol and morphine, and it is also obvious that Demerol was almost always as active spasmogenically as morphine in each patient. The work of Gaensler and his associates is an important contribution to the pharmacology of Demerol in man, and, although it confirms the results obtained in the dog prepared with the Thiry-Vella loop, it comes somewhat as a surprise in view of the general spasmolytic rather than spasmogenic

action of Demerol in the gastrointestinal tract of man (except in the duodenum and jejunum).

Because of the work of Gaensler and his associates, we were eager to determine the effects of Demerol on one portion of the colon which had not been previously studied, namely, the rectum. The effects of morphine on the entire colon had been well known,¹¹ as illustrated in FIGURE 15.

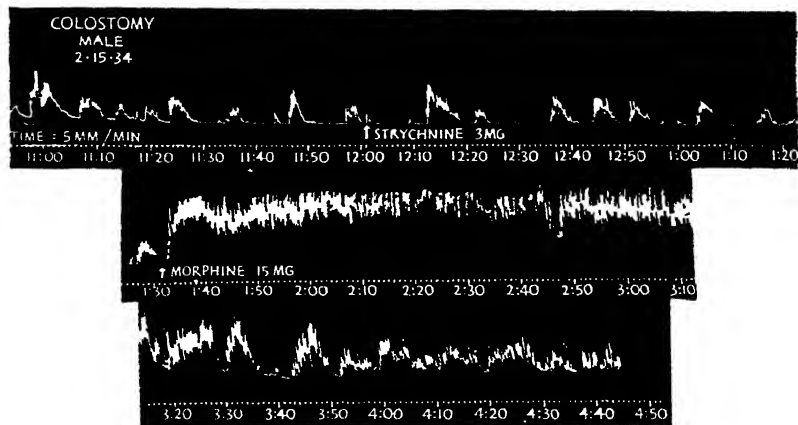


FIGURE 15. Colostomy (male).

The spasmogenic action of morphine undoubtedly accounts for the marked constipation observed in many patients postoperatively and in addicts, if the same action prevailed following Demerol, this would obviously be a decided deterrent to its broad clinical usage. Clinical experience over several years had failed to support the idea that Demerol was constipating like morphine. The kymographic evidence was not at hand and we therefore approached Dr. Frederick R. Steggerda¹² of the Department of Physiology at the University of Illinois in Urbana, to make such studies with his specially designed rectal insufflation tube technique. FIGURE 16 indicates the relaxing effect of Demerol on the distal colon of three different subjects. In no instance was Demerol spasmogenic as it is in the duodenum and jejunum. This failure of spasmogenicity prevailed despite marked side effects in one patient which proceeded to the point of fainting. One notes that, in each of the three subjects studied, a dose of 100 mg. of Demerol intramuscularly elicited either no side effects, moderate side effects or severe side effects in the respective subjects, a finding supported amply by wide clinical usage.

The failure of Demerol to act spasmogenically like morphine in the colon of man is further emphasized by FIGURE 17, which presents a kymographic tracing obtained from a young man with a colostomy. This patient, at the Detroit Receiving Hospital, had received Demerol in a dosage of 250 mg. day for four days prior to this experiment. One notes

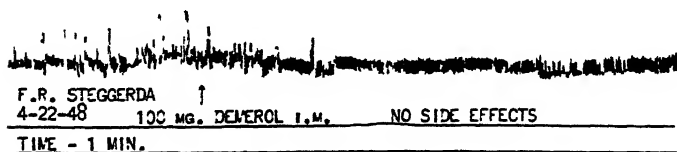
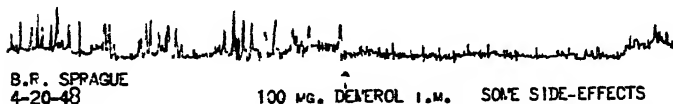
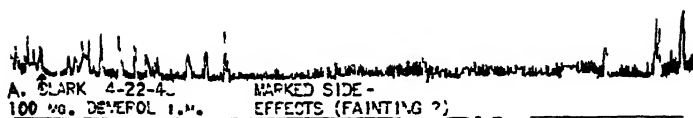


FIGURE 16. Demerol in man.

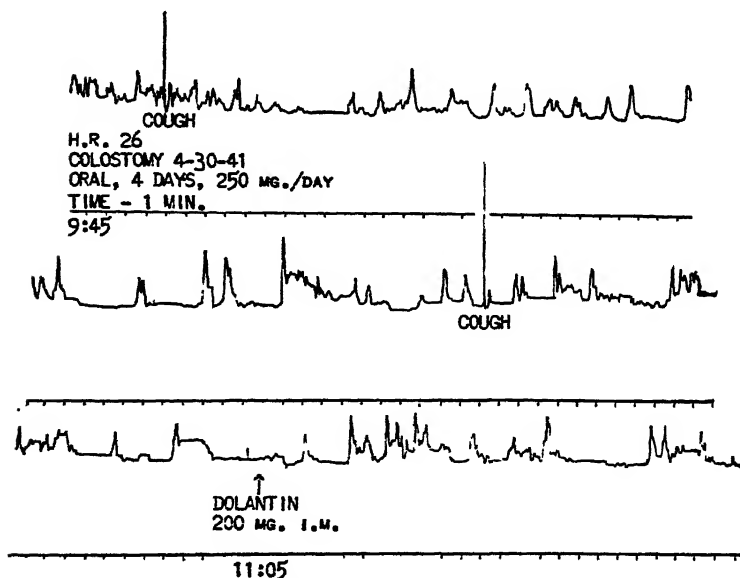


FIGURE 17. Demerol in man.

that the colonic tracing is quite uniform and presented no variations from several previously obtained control tracings even after 200 mg. of

Demerol (Dolantin) had been injected intramuscularly. It is quite evident why Demerol is to be preferred clinically in those patients who react prominently by way of a spasmogenic effect to morphine following either acute or chronic administration of the drug as indicated.

Analgesia

The analgesic power of Demerol has been tested by various means. As a rule, one of the important test objects has been the white rat. FIGURE 18 is taken from the work of Dr. Tainter and his associates¹³ at the Ster-

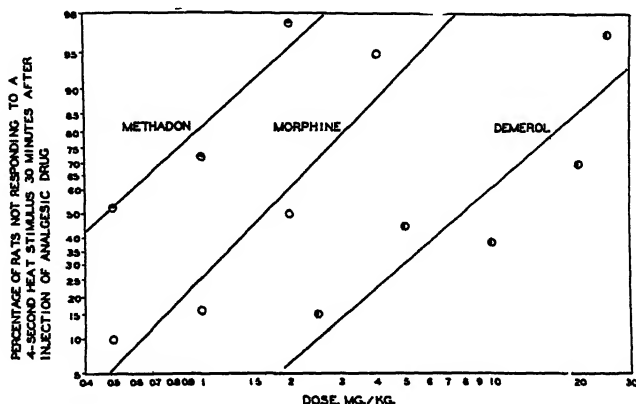


FIGURE 18. Comparison of the analgesic power of Demerol, methadone, and morphine, as tested by the proportion of rats responding to a 4-second heat stimulus to the skin 30 minutes after subcutaneous injection of graded doses of the three analgesics. (From Tainter.¹³)

ling-Winthrop Research Institute and indicates the comparative analgesic potencies of Demerol, methadone, and morphine. It is noted that the various analgesic agents are effective in 50 per cent of the animals tested in the following milligram per kilogram dosages: methadone 0.48, morphine 1.65, and Demerol 7.4; hence, methadone was 3.4 times more potent than morphine, whereas it was 15 times stronger than Demerol. By this standard and test as employed, methadone was the most potent analgesic of the three compared.

The analgesic power of Demerol had been compared earlier with that of morphine and codeine in human subjects using the Wolff-Hardy-Goodell technique, by various investigators and particularly by Dr. Barlow, formerly of the Sterling-Winthrop Institute. FIGURE 19 depicts the comparative action of two doses of Demerol in terms of codeine and morphine analgesia. It is noted that Demerol ranks between codeine and morphine, and although it is not quite as potent an analgesic agent as morphine it is definitely superior to codeine in the ordinary doses employed.

It was our privilege to have participated in the study of Demerol in clinical subjects¹⁴ while at Wayne University, and I wish gratefully to

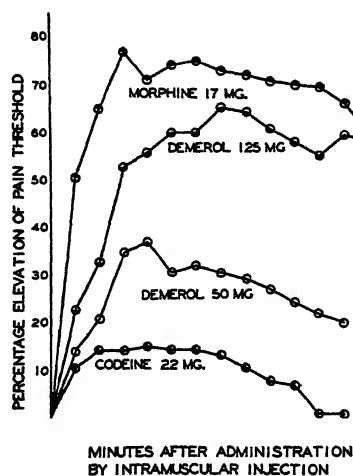


FIGURE 19. Comparison of analgesic power of morphine, Demerol, and codeine in human subjects when tested by the pain threshold of the skin to heat stimuli. The curves are the averages of 16 experiments. (From TAINTER.¹³)

acknowledge the fine cooperation of our former associates, Dr. Paul Noth and Dr. Hans Hecht. Dr. Noth, of the Department of Medicine at Wayne University, conducted his studies at the Detroit Receiving Hospital and Dr. Hecht, now of the University of Utah but formerly of the same Department, made his studies at the William J. Seymour Hospital, Eloise, Michigan. It was a genuine pleasure to have worked with both of them.

The appraisal of the analgesic effect of Demerol was made in terms of the degree of relief obtained, that is, either complete or partial relief, or failures expressed as no relief. It is of interest to note the varying types of pain which were studied. As one scans TABLE 2, one observes that pain

TABLE 2
ANALGESIC EFFECT OF DEMEROL

<i>Types of pain</i>	<i>Results</i>			
	<i>No. of cases</i>	<i>Complete relief</i>	<i>Partial relief</i>	<i>No relief</i>
Arthritic	9	6	2	1
Gastrointestinal	22	13	6	1
Biliary tract	5	3	1	1
Renal	6	6		
Osseous and/or nerve	26	14	8	4
Headache	18	13	2	3
Pleural	12	8	2	2
Cardiac	10	7	2	1
Miscellaneous	15	9	4	2
Total	123*	81 (65.8%)	27 (22.0%)	15 (12.2%)

* Five patients had 2 types of pain.

associated with spasm of a hollow viscus seemed to lend itself most readily to Demerol therapy. The two failures noted under gastrointestinal and biliary tract types of pain could have been due to the spasmogenic effect of Demerol in the upper part of the digestive tract. Despite the fact that Demerol is an analgesic agent centrally, the functional pathology prevailing at the site or source of pain in the upper digestive tract might have been too severe (increased spasm!) to be adequately covered by the central analgesic action of Demerol. An attempt was made to break down the analgesic appraisal into three groups, which, in reference to the degrees of pain, were classified as very severe, severe, and moderately severe. The degree of relief was expressed, as in TABLE 3, in terms of com-

TABLE 3
ANALGESIC EFFECT OF DEMEROL
(127 instances in 81 patients)

Degree of pain	Results					
	Complete relief		Partial relief		No relief	
	No. instances	Group %	No. instances	Group %	No. instances	Group %
Group 1 Very severe (44 instances)	26	59.1%	8	18.2%	10	22.7%
Group 2 Severe (50 instances)	29	58.0%	14	28.0%	7	14.0%
Group 3 Moderately severe (33 instances)	25	75.7%	6	18.2%	2	6.1%
Combined groups (127 instances)	80	63.0%	28	22.0%	19	15.0%

plete and partial relief, or failure expressed as no relief. One observes that the analgesic effect of Demerol is 85 per cent, which closely approximates that set forth in TABLE 2, namely, about 87 per cent.

Sedation

Another important pharmacodynamic action of Demerol is that of sedation. Although this is not as prominent with Demerol as that experienced with morphine, it can be made use of in certain patients. TABLE 4 indicates the degree of sedation observed, with or without sleep, in three varying degrees of pain previously described, namely, very severe, severe, and moderately severe. One observes that sleep was induced in approximately 50 per cent of the patients studied. This might well have been due to the fact that pain in many of these patients, especially in group 1, had been so severe that considerable insomnia was endured by the patient because of attending pain. The analgesic action of Demerol, resulting in obtunding of the pain, might have permitted sleep in the exhausted patient so that sedation, although apparent, might have been

TABLE 4
SEDATIVE EFFECT OF DEMEROL IN PRESENCE OF PAIN
(99 instances in 81 patients)

<i>Degree of pain</i>	<i>Results</i>					
	<i>Sleep induced</i>		<i>Sedation without sleep</i>		<i>No effect</i>	
	<i>No. instances</i>	<i>Group %</i>	<i>No. instances</i>	<i>Group %</i>	<i>No. instances</i>	<i>Group %</i>
Group 1 Very severe (33 instances)	18	54.5%	6	18.2%	9	27.3%
Group 2 Severe (37 instances)	15	40.5%	15	40.5%	7	19.0%
Group 3 Moderately severe (29 instances)	17	58.6%	9	31.0%	3	10.4%
Combined groups (99 instances)	50	50.5%	30	30.3%	19	19.2%

gained indirectly by relieving pain as the cause of sleeplessness, rather than by a direct sedative property of the analgesic agent. Sedation without sleep was observed in approximately 30 per cent, but on the other hand 19 per cent failed to demonstrate any sedative effect. This is in decided contrast with morphine, which invariably produces sedation in moderate or severe degrees, depending on the dosage employed.

Side Effects

Varying types of side reactions or undesirable sequelae were observed in several of our patients. These were of varying degrees and intensities, and the types and severities are enumerated in TABLE 5. In seven cases of those studied, it was necessary to withdraw or alter the medication because of the severity of the side reactions associated with Demerol therapy. Some patients experienced subjective vertigo on very small doses of the drug, namely, 25 mg. as a total dose, whereas others tolerated it very well without the least side reaction on doses as high as 300-400 mg. or more daily. Induration of the tissues at the site of injection seemed to be associated chiefly with the technique of administration. In other words, intramuscular injections were well tolerated but superficial subcutaneous injections frequently invoked an induration reaction. Dryness of the mouth was to be anticipated because of the weak anticholinergic action which Demerol exerts. Euphoria is of special note because of the capacity of Demerol to produce physiological dependence upon it. In this respect, it is akin to morphine and alcohol, but in varying degrees. Further scanning of TABLE 5 indicates the usual types of side reactions associated with many new drugs, but special note should be taken of that single instance in which vascular collapse with bronchial spasms

TABLE 5
SIDE EFFECTS OF DEMEROL
(Experienced by 40 of 116 patients)

<i>Symptoms</i>	<i>No. of instances*</i>
Subjective vertigo	10 (1)
Induration of tissues at site of injection	10
Dryness of mouth	10
Euphoria	8
Nausea and vomiting	5 (3)
Paresthesias	4
Nausea	3
Sweating	2
Muscular relaxation	2
Palpitation and increased dyspnea	1 (1)
Vascular collapse with bronchial spasm	1 (1)
Insomnia and restlessness	1 (1)
Transient dimness of vision, pain at site of injection, drug eruption (?), delusions, urticaria—each	1
<i>Total</i>	<u>62 (7)</u>

* Numbers in parenthesis indicate cases in which side effects necessitated stopping Demerol.

was observed. This calls to mind the important pioneer studies of Gruber and his associates dealing with acute hypotension following administration of Demerol intravenously in the dog. There have been other reports of this type appearing in the literature since this announcement, but it should be stressed that Demerol can be given safely and with a fair degree of impunity if it is injected slowly in adequate fluid volume.

The frequency and varieties of side effects from Demerol have been thoroughly studied by Dr. Batterman.⁸ It is his conclusion that side reactions are much more severe and frequent in the ambulatory than in the hospitalized patient. Furthermore, it seems to be the rule that medication of this type when administered orally is more of an offender than when it is administered parenterally. The same conclusion has been drawn by Dr. Batterman from his experiences with methadone, as indicated in TABLE 6.

The matter of addiction to Demerol has received considerable attention in both the foreign and American literature, and since this topic will be very adequately discussed by several authorities participating in this conference we shall refer to it only very briefly at this time. We have observed physiological dependence to Demerol in our series of patients, but it is the author's impression that Demerol, at the time our studies were made, was not necessarily primarily addicting since most physiological dependence, as observed in our patients, occurred chiefly in those subjects who had been addicted previously to morphine or other agents of this type.¹⁴ Although addiction to Demerol can be produced, it is apt to occur not only less frequently and less intensively but can also be more readily corrected than addiction to morphine.

The important actions and uses of Demerol as compared with those

TABLE 6

PERCENTAGE FREQUENCY AND VARIETIES OF SIDE EFFECTS FROM METHADONE*

Kind of side effect	Hospitalized patients		Ambulatory patients by oral route
	Parenteral route	Oral route	
	%	%	%
Dizziness	11.1	20.0	70.0
Nausea	4.4	15.1	40.0
Vomiting	3.3	4.6	18.6
Diaphoresis	3.5	9.2	18.6
Epigastric pain		1.5	
Dryness of mouth	1.1	4.6	3.3
Anorexia	2.2	9.2	10.0
Pruritus		1.5	6.6
Weakness		12.3	23.3
Headache	2.2	3.2	10.0
Drowsiness	3.5	13.8	33.3
Grogginess	3.3	4.6	6.6
Miosis	2.2	4.6	
Urinary retention		1.5	
Visual disturbances		4.6	6.6
Mental confusion and toxic psychosis	1.1	9.2	
Numbness			6.6

* From BATTERMAN & OSHLAG.²⁴

of morphine can best be summarized in TABLE 7, which it was our good fortune to receive from Dr. Goodman and Dr. Gilman as it is planned for their revised edition of *The Pharmacological Basis of Therapeutics* which we trust will soon be forthcoming.

Some of these points are worth special attention. For example, respiration is never depressed by isonipecaine (Demerol) but is uniformly depressed by morphine, making the former the drug of choice by many obstetricians. Not only is the mother's respiration maintained quite readily at status quo but there is seldom, if ever, any significant degree of respiratory embarrassment of the newborn which is quite the contrary in terms of morphine.

Smooth muscle of the GI tract is uniformly rendered spastic with morphine, whereas it is either unaffected or relaxed by Demerol with the singular exception of the duodenum, as previously reported in reference to the work of Gaensler and his associates.

The comparative degrees of euphoria and tolerance probably account considerably for the differing likelihood of addiction to Demerol and morphine, the drug producing the greater degree of euphoria and tolerance likewise being more likely to produce addiction, as is the case with morphine.

The prominent symptom of overdosage associated with the action of these drugs on the central nervous system is that of excitement resulting from Demerol and depression from morphine, effects in direct apposition.

As a preanesthetic medicament, Demerol is being given wide usage,

TABLE 7
A COMPARISON OF ISONIPECINE* AND MORPHINE

Source	<i>Isonipeccine</i> Synthetic piperidine deriva- tive	<i>Morphine</i> Natural alkaloid of opium
Discovery	1939, Eisleb & Schaumann	1803, Sertürner
Locus of C.N.S. action	Cortex and diencephalon	All segments of cerebro- spinal axis
Type of C.N.S. action	Depression and excitation	Depression and excitation
Ratio of sedation: analgesia	Low	High
Analgesic potency (weight basis)	1	10
Respiration	Rarely depressed	Uniformly depressed
Cough reflex	Not depressed	Depressed
Pupils	Usually unaltered	Characteristic miosis
Corneal reflex	Markedly obtunded	Rarely affected
Smooth muscle	Spasmolysis	Spasm
Secretions	Inhibited moderately	Inhibited slightly
Cardiovascular effects	Benign, occasionally syn- cope	Benign
Euphoria	Inconstant, moderate	Frequent, marked
Addiction liability	Moderate, rarely serious	Marked, frequently serious
Tolerance	Moderate, rare	Marked common
Untoward side effects	Many, usually transient, not serious	Many persistent, trouble- some
Dosage	30 to 900 mg.	5 to 20 mg.
Route of administration	Oral, intramuscular	All routes
Duration of analgesia	3 hours +	4 hours +
Overdosage	Excitement	Depression
Uses:		
Relief of pain	Satisfactory	Satisfactory
Antiperistaltic	Not useful	Useful
Cough	Not useful	Useful
Sedation and sleep	Occasionally useful	Often useful
Antispasmodic	Satisfactory	Contraindicated
Pranesthetic medication	Satisfactory	Satisfactory
Labor	Satisfactory	Care!

* Demerol.

and the same statement can be made for its selection in the field of obstetrics where it has become a very valuable drug primarily because of its lack of respiratory effects in efficient analgesic dosage.

As stated earlier, limitations of time necessitated severely restricting material for this presentation. There are many other extremely valuable and interesting reports available which deal with still other pharmacological features of Demerol, but we regret that we can only refer to some of these as they pertain to studies dealing with electroencephalography,¹⁵ local anesthesia,¹⁶ fate,¹⁷⁻¹⁹ and general metabolic effects.⁸

Derivatives of Demerol

Numerous attempts have been made to prepare various derivatives of

Demerol. MacDonald and his associates²⁰ have prepared an isomer of Demerol, known as iso-pethidine (FIGURE 20), which was studied by vari-

DISCUSSION

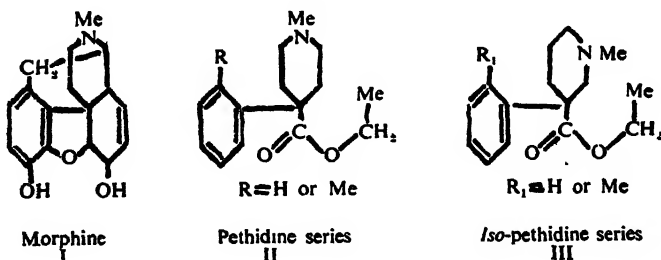


FIGURE 20. (From MACDONALD *et al.*²⁰)

ous techniques. Glazebrook and Branwood²¹ used as their criterion the ability of iso-pethidine to depress pain afflicted by a sharp point under varying degrees of pressure as produced by inflation of a modified blood pressure unit (FIGURE 21). Analgesia was calibrated in terms of the excess

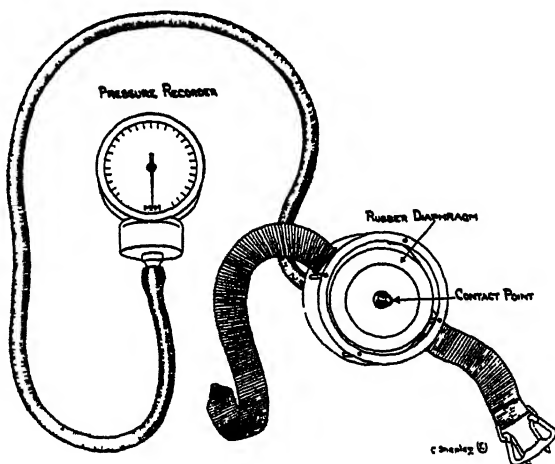


FIGURE 21. Device for measuring pain threshold. (From GLAZEBROOK & BRANWOOD.²¹)

pressure required to inflict pain after the administration of iso-pethidine, pethidine, and coco-tabs (combination of aspirin, Phenacetin, and codeine). FIGURE 22 indicates that pethidine is still somewhat stronger as an analgesic than is its isomer, iso-pethidine.

Other derivatives of Demerol have been prepared and are illustrated in FIGURE 23. Some of these compounds are apparently more effective analgesics when appraised in various types of experimental animals, and present clinical trials indicate definite promise in terms of specific advantages. Further clinical appraisal is necessary, however, before final

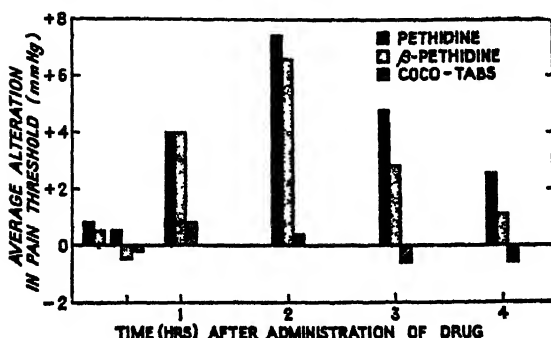


FIGURE 22. Average amount of alteration in pain threshold measured in mm. Hg in each group of 20 cases. No coco-tab column is shown at the quarter-hour mark, because no difference in pain threshold was recorded at this time. (From GLAZEBROOK & BRANWOOD.¹¹)

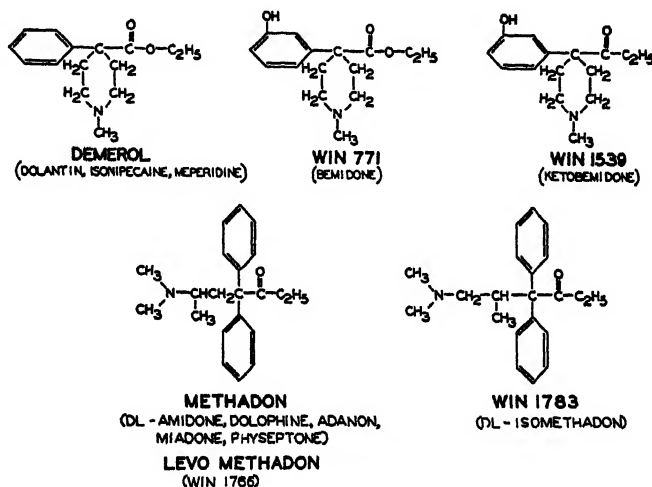


FIGURE 23. Structural formulae and synonymous names of some of the new synthetic analgesics. (From TAINTER.¹⁴)

conclusions can be drawn concerning comparative desirability of these Demerol or pethidine derivatives as prepared in the Winthrop laboratories.

Another important and interesting series of piperidine derivatives (prepared by Lee and his associates²²) has been studied by Foster and Carman,²³ who draw the following important conclusions concerning the relation of pharmacodynamic activity to chemical manipulation:

1. The 4-phenyl radical is essential, *e.g.*,

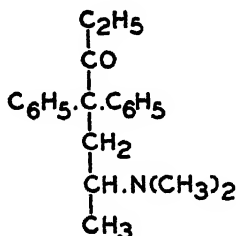
5. GRUBER, C. M., E. R. HART, & C. M. GRUBER, JR. 1941. Pharmacology and toxicology of ethyl ester of 1-methyl-4-phenyl-piperidine-4-carboxylic acid (Demerol). *J. Pharm. & Exp. Therap.* **73**: 319.
6. YONKMAN, F. F., P. H. NOIH, & H. H. HECHT. 1944. Demerol: a new synthetic analgesic spasmolytic and sedative agent. I. Pharmacologic studies. *Ann. Int. Med.* **21**: 7.
7. PLANT, O. H., & G. H. MILLER. 1926. *J. Pharm. & Exp. Therap.* **27**: 361.
8. BATTERMAN, R. C. 1943. The clinical effectiveness and safety of a new synthetic analgesic drug, Demerol. *Arch. Int. Med.* **71**: 343.
9. GAENSLER, E. A., J. M. MCGOWAN, & F. F. HENDERSON. 1948. A comparative study of the action of Demerol and opium alkaloids in relation to biliary spasm. *Surgery* **23**: 211.
10. MCGOWAN, J. M., W. BUTSCH, & W. WALLERS. 1936. Pressure in the common bile duct of man: its relation to pain following cholecystectomy. *J. A. M. A.* **106**: 2227.
11. YONKMAN, F. F., J. M. HIEBERT, & H. SINGH. 1936. Morphine and intestinal activity. *New England J. Med.* **214**: 507.
12. STEGGERDA, F. R. Personal communication.
13. TAINTER, M. L. 1948. Estudio comparativo de algunas acciones del Demerol y del methadon. *Bol. Asoc. Med. Puerto Rico* **40**: 1.
TAINTER, M. L., & O. H. BUCHANAN. 1948. A comparison of certain actions of Demerol and methadon. *California Med.* (In press.)
14. NOIH, P. H., H. H. HECHT, & F. F. YONKMAN. 1944. Demerol: a new synthetic analgesic, spasmolytic and sedative agent. II. Clinical studies. *Ann. Int. Med.* **21**: 17.
15. ANDREWS, H. L. 1942. Cortical effects of Demerol. *J. Pharm. & Exp. Therap.* **76**: 89.
16. WAY, E. L. 1946. Studies of local anesthetic properties of Isonipecaine. *J. Am. Pharm. A. (Sci. Ed.)* **35**: 44.
17. BERNHEIM, F., & M. L. C. BERNHEIM. 1943. The hydrolysis of Demerol by liver *in vitro*. *J. Pharm. & Exp. Therap.* **85**: 74.
18. LEHMAN, R. A., & T. AITKEN. 1943. Determination of Demerol in urine with preliminary observations of its excretion in man. *J. Lab. & Clin. Med.* **28**: 787.
19. OBERST, F. W. 1943. Method for determination of Demerol in urine and results of its application. *J. Pharm. & Exp. Therap.* **79**: 10.
20. MACDONALD, A. D., *et al.* 1946. Analgesic action of pethidine derivatives and related compounds. *Brit. J. Pharm.* **1**: 12.
21. GLAZEBROOK, A. J., & A. W. BRANWOOD. 1945. Clinical trials of β -pethidine. *Lancet* **2**: 528.
22. LEE, J., L. BURGER, A. ZIERING, & S. C. HILNEMAN. 1946. Jubilee Volume Emil Barrell: 264. Basel.
23. FOSTER, R. H. K., & A. J. CARMAN. 1947. Studies in analgesia. Piperidine derivatives with morphine-like activity. *J. Pharm. & Exp. Therap.* **91**: 193.
24. BATTERMAN, R. C., & A. M. OSHLAG. 1948. *Fed. Proc.* **7**: 206; *J. A. M. A.* (In press.)

PHARMACOLOGY OF METHADONE AND RELATED COMPOUNDS

BY K. K. CHEN

The Lilly Research Laboratories, Indianapolis, Indiana

INTRODUCTION. The development of methadone, in the United States, as an analgesic drug in rivalry with morphine is one of the medical advances resulting from World War II. It was the State Department that sent the Technical Industrial Intelligence Committee, composed of Kleiderer, Rice, Conquest, and Williams, to investigate the research activities of I. G. Farbenindustrie at Hoechst-am-Main.¹ Among the 25 projects examined, the one on analgesics appears to be the most outstanding. It is an offshoot of Demerol research, but deals with a different class of compounds. One member of the series having the chemical formula of 6-dimethylamino-4,4-diphenyl-3-heptanone, shown below, is 5 to 10 times as active as Demerol:



The Council on Pharmacy and Chemistry of the American Medical Association has adopted the non-proprietary name of methadone.² Manufacturers in the United States call it "Dolophine" (Methadon, Lilly) and "Adanon" (Winthrop). In England, it is known as miadone. The I. G. serial number for this compound was 10820, and the code-name for clinical testing in Germany, amidon. The latter was derived from pyramidon by the omission of the first syllable. According to direct information from Germany, the term amidon would not have been used for marketing of this product. Unfortunately, in the United States, the word amidone, with an "e" at the end, is becoming increasingly common in scientific literature. This name is not very desirable, because confusion may arise from "*Tamidon*," the French word for starch. The alcohol corresponding to methadone has been referred to as "amidol"; but the latter has already been used to designate a photographic developer, diamino-phenol. It would seem better if the name amidone could be dropped entirely.

One of the questions which come to anybody's mind is, why the Germans did not exploit the possibilities of methadone in military and civilian medicine during the War, especially in view of the morphine short-

age. Through correspondence with a former I. G. employee, it was learned that both the German firm and the military authorities discredited the product because of its side effects. It was probable that the doses they employed were larger than necessary and thus resulted in untoward symptoms.

Action on the Central Nervous System. Much more information is now available regarding the pharmacology, toxicity, clinical uses, and addiction liabilities of methadone, as a result of intensive investigations carried out in various laboratories in this country. Like morphine, it has a profound action on the central nervous system.³ It raises pain threshold by the depression of the somesthetic area of the brain. In rats, a dose of 1 mg. per kg. injected intraperitoneally is sufficient to nullify the pain from pinching of the tail, according to the Haffner technique.⁴ Similarly, in dogs, the same dose injected subcutaneously elevates the thermal threshold to cause a skin-twitch of the back, measured by the Andrews modification⁵ of the Hardy-Wolff-Goodell method.⁶ In man, by the latter procedure, methadone in the dose of 2.5 mg. given by mouth raises the thermal pain threshold of the forehead. When compared with morphine sulfate, methadone, in form of hydrochloride, is twice as active analgesically, weight for weight.⁷ This is true in all three species of animals—rats, dogs, and man. The relief of pain by methadone in clinical cases has been amply proved.⁸⁻¹⁴

The new drug produces sedation^{8, 9, 10, 12} by the depression of the sensory areas of the brain cortex. This can be demonstrated in both cold-blooded¹⁵ and warm-blooded animals. The amount required to bring about sedation is substantially greater than that for analgesia. In this respect, methadone is relatively less effective than morphine. As an illustration, methadone HCl in a 2.5-7.5-mg. dose relieves pain without sedation in human subjects, while morphine sulfate in a 15-mg. dose also relieves pain, but is accompanied by lessened activity, mental sluggishness, and a desire to sleep.

Methadone in rather large doses depresses respiration, presumably by direct action on the respiratory center.³ This is manifested in anesthetized dogs by a decrease of respiratory rate, and, frequently, amplitude, as well as by a decrease in the volume of expired air. Fall of respiratory rate has been observed in man with doses of 10-30 mg.^{16, 17}

In anesthetized dogs, an intravenous injection of methadone is also followed by a decrease of heart rate.^{3, 7} The decrease may be prevented by the section of vagi, or by intravenous injection of atropine. The action is, therefore, attributable to the stimulation of the vagal center. The same effect is exerted on the intestines.^{3, 18, 19} Methadone, injected intravenously, stimulates peristaltic movements of intestines *in situ* of non-anesthetized dogs, which is antagonized by atropine. Since methadone uniformly inhibits the isolated intestinal movements, its action here must be central in origin—namely, on the vagal center.

One of the side effects in human subjects from an overdose of methadone is vomiting, preceded by nausea.^{3, 10, 12} It appears more frequently in individuals who move around than in those who remain in bed. The emetic action of methadone is, most likely, due to the stimulation of the vomiting center, and not to local irritation of the gastric mucosa, since vomiting occurs after parenteral administration of the drug. It is interesting that such animals as pigeons, dogs, and monkeys, which are very sensitive to emetic drugs, do not vomit following the administration of methadone, in any quantity. So far as is known to date, man is the only mammal susceptible to the emetic action of this new analgesic.

Methadone, when injected subcutaneously into mice, causes continuous erection of the tail, in exactly the same manner as morphine.³ An example is shown in FIGURE 1. This phenomenon for morphine has

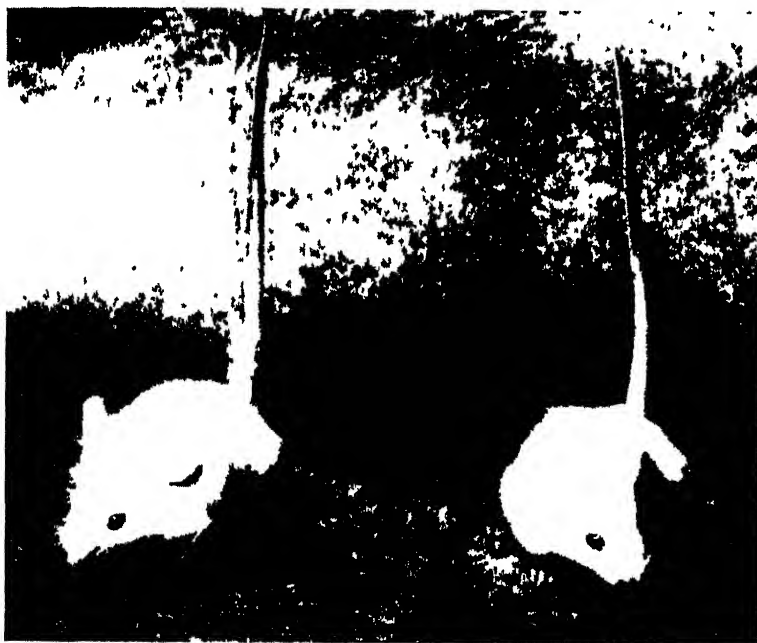


FIGURE 1. Action of methadone and morphine on the mouse tail. The animal on the left-hand side received methadone in the dose of 2 mg. per kg., and that on the right-hand side, morphine in the dose of 5 mg. per kg.

been interpreted by Heinekamp²⁰ as due to the stimulation of the spinal cord. It is conceivable that methadone may follow the same pattern of action. Leopard frogs, African clawed frogs, or turtles, after receiving a lethal or sublethal dose of methadone, undergo a period of narcosis, followed by hyperirritability.¹⁵ During the latter stage, a gentle tapping results in a quick twitch of all muscles. Pithing, but not destruction of the medulla, abolishes all muscular movements. Mild as it may be, the re-

action may be construed as evidence of stimulation of the spinal cord. On the other hand, Wikler and his colleagues²¹ and Leimdorfer²² obtained negligible evidence in favor of cord action in cats and dogs.

The action of methadone is, therefore, mixed and complicated. It depresses the sensory areas of the brain cortex in the form of analgesia and sedation; and it depresses the respiratory center of the medulla. On the other hand, it stimulates the vomiting center of the medulla in man; it stimulates the vagal center controlling the heart and intestines; and it has a questionable action on the spinal cord.

Action on Smooth-Muscle Organs. The I. G. workers recorded¹ that methadone relaxed isolated guinea pig's intestines and released their spasm induced by musculotropic substances such as histamine and barium chloride, or by cholinergic substances such as carbamyl choline. Similar results of the same experiments with Demerol^{23, 24} made them look upon the entire class of compounds as antispasmodics. The analgesic property was an incidental finding. The inhibitory action of methadone on isolated intestines of rabbits and guinea pigs was also reported by American investigators.^{3, 25, 26}

Gruber and his associates²⁷ had already shown that Demerol has a stimulating action on intestines *in situ*. The same happens to methadone. It has been repeatedly pointed out that, in trained non-anesthetized dogs, the action of methadone on ileum or jejunum *in situ* is purely excitatory.^{3, 19} It appears, then, that in the intact animal methadone stimulates intestines by acting on the vagal center, as mentioned above; but if the central connection is severed, as in the case of isolated intestines, the drug is primarily inhibitory by direct action on smooth muscles.

Regarding the isolated uterus, the action of methadone appears to depend on its initial activity. If it is in rhythmic movements, the drug causes relaxation; but if it is inactive at the start, the drug usually produces contractions.^{26, 28} The isolated rat's and hamster's uteri are always rhythmically active, and they respond to methadone by relaxation. The isolated rabbit's uterus may be active or inactive, and its response is usually opposite to its own state of motility. In case of a spasm induced by mecholyl or ergonovine, methadone consistently shows a relaxing effect. The rabbit's uterus *in situ* is frequently inhibited by methadone.

The ureter is another smooth-muscle organ, the reaction of which is of more than academic interest, because the drug may be thought of in renal colic. In anesthetized dogs, Dr. Lee of our laboratory²⁹ observed relaxation of ureteral movements recorded by the Tratner method.³⁰

The action of methadone on smooth-muscle organs is, therefore, also complicated. In the intact animal, it appears to stimulate intestines, but to inhibit the uterus and the ureter. When isolated from the body and immersed in Tyrode's or Ringer's solution, the intestines uniformly respond by relaxation, and the uterus moves in the opposite direction in accordance with its own activity.

Local Anesthesia, Hyperglycemia, Hypothermia, Absorption, and Elimination. Everett³¹ first reported that methadone had a local anesthetic action on the rabbit's cornea, the duration of which equaled that of cocaine. He emphasized its irritating properties and advised against its practical use. Unpublished data of our own laboratory showed that 0.1 cc. of a 1 per cent solution of methadone injected intracutaneously into guinea pigs produced local anesthesia lasting for an average of 85 minutes. The same solution, when instilled in the rabbits' eyes, resulted in corneal anesthesia for an average of 13 minutes, with marked chemical inflammation. Barring its clinical application, the observation indicates that methadone temporarily paralyzes peripheral, sensory nerve fibers—an action not shared by morphine.

Rise of blood sugar in dogs and rabbits following parenteral administration of methadone has been pointed out by Wikler, Haag, Phatak, and their respective associates.^{21, 32-34} Our own studies on rabbits by intravenous injection of the drug fully confirm their results, as exemplified in FIGURE 2.

Fall of rectal temperature in dogs has been observed by Wikler and

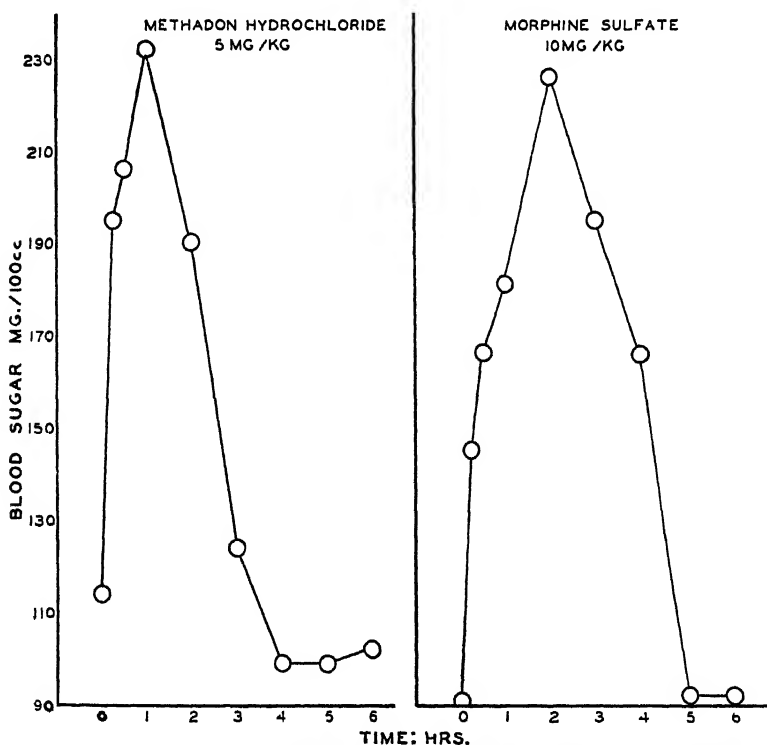


FIGURE 2. Hyperglycemia in rabbits caused by methadone and morphine. Both drugs were injected intravenously. The two animals were from the same stock, and of about the same weight.

his co-workers.²¹ The mechanism of hypothermia as well as hyperglycemia is, presumably, the same as that which occurs with morphine.

Methadone is easily absorbed from almost any common route of administration. The question is—what does the body do with it? By the bromthymol blue method, Scott and Chen² recovered 20–35 per cent of unchanged methadone in the urine. Cronheim and Ware,³⁵ using brom-cresol purple, found only a 6–13 per cent excretion of the administered drug. In any event, a considerable amount is metabolized in the body, the process of which is as yet unknown. The latter may be elucidated by the radioisotope technique, like those of many other substances. Work along this line is now in progress.

Toxicity. There is an accumulation of toxicity data published from various laboratories. By a single injection into experimental animals, the lethal doses vary considerably from species to species. As shown in TABLE I, methadone is more toxic than morphine, in the ratios of 7:1 up

TABLE I
COMPARISON OF ACUTE TOXICITY BETWEEN METHADONE AND MORPHINE
BY SUBCUTANEOUS INJECTION

Animal	Median lethal dose \pm standard error: mg. per kg.		Ratio of morphine to methadone
	Methadone HCl	Morphine sulfate	
Leopard frog	102 \pm 20	908 \pm 110	1:9
African clawed frog	55.5 \pm 5.3	677 \pm 68	1:12
Turtle	31.8 \pm 2.2	253 \pm 51	1:8
Guinea pig	54.4 \pm 3.6	391 \pm 25	1:7
Mouse	33.7 \pm 5.4	311 \pm 53	1:9
Rat	12.4 \pm 2.5	229 \pm 46	1:18

to 18:1 in different animals. It is particularly toxic to the rat. Woods, Wyngaarden, and Seever^{36, 37} reported that the lethal dose of methadone by subcutaneous injection in the rhesus monkey lay between 10 and 20 mg. per kg.

Death in warm-blooded animals is due to respiratory failure. Although some vasoconstrictors are efficient stimulants of respiration during methadone depression,¹⁸ it is uncertain that they will save lives from lethal doses.

Toxicity studies by prolonged administration of methadone in animals were carried out in Haag's laboratory^{32, 33} and our own.³ In short, the drug does not cause any somatic damage which can be detected grossly or microscopically. Tolerance to the analgesic and sedative action of methadone in dogs and man has been demonstrated by the scientists of the U. S. Public Health Service at Lexington, Kentucky.^{12, 38, 39} Acute vascular tolerance to methadone in anesthetized dogs has been reported by Shideman and Johnson.⁴⁰

The lethal dose in man is unknown. Isbell and his associates,¹² in their exhaustive investigations, employed a dosage level of 200 mg. 4 times

daily in one case, and 150 mg. 4 times daily in another. In most patients, the dose should be adjusted from 2.5–10 mg. Very rarely does it need to be increased by 15–20 mg.

Similarities to and Differences from Morphine. It is interesting that such a simple compound, a keto-diphenyl-tertiary amine, could exert actions on organisms so similar to morphine, an alkaloid of the phenanthrene series. The similarities can be enumerated as follows—analgesia, respiratory depression, emesis in man, hyperglycemia, hypothermia, tolerance, and addiction in dogs and man,^{12, 38} as well as erection of the mouse tail. Scott and Chen⁸ reported that methadone, like morphine, causes excitement in cats. In rabbits, both morphine and methadone decrease the propulsive activity of intestines, as shown by Karr,⁴¹ thus accounting for their constipating effect.

There are differences, however, between the two drugs. The depressor action of morphine is much greater than that of methadone. FIGURE 3

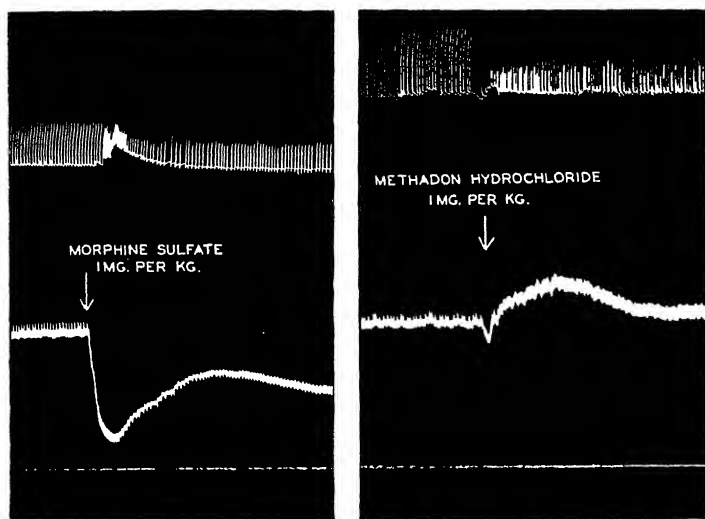


FIGURE 3. Action of methadone and morphine on blood pressure. Records were made from two anesthetized dogs, one with each drug. Tracings from top to bottom are respiratory movements, carotid blood pressure, and base line. All injections were made intravenously.

shows that a dose of 1 mg. per kg. results in a much greater fall of blood pressure with morphine than with methadone in two separate anesthetized dogs. Other differences are listed in TABLE 2. It has already been mentioned that methadone has a local anesthetic action.³¹ Unpublished results of this laboratory indicate that methadone protects guinea pigs from histamine aerosol, and releases histamine spasm of their isolated intestines. These antihistaminic effects are weak as compared with benadryl, but they are, nevertheless, absent with morphine. It has been

TABLE 2
DIFFERENCES BETWEEN METHADONE AND MORPHINE

<i>Action</i>	<i>Methadone</i>	<i>Morphine</i>
Depressor action	slight	marked
Local anesthesia	+	—
Antihistaminic	+	—
Convulsions in amphibians	—	+
Sedation in small clinical doses	—	+
Inhibition of hexokinase	+	—
Age effect on toxicity	—	+
Euphoria in non-addicts	?	+
Physical dependence	slight	intense
Addiction in monkeys	—	+

pointed out above that cold-blooded animals do not develop convulsions with lethal or sublethal doses of methadone, and that human subjects do not experience sedation with small therapeutic doses of the new drug—in contrast with morphine. Greig⁴² demonstrated that methadone inhibits the glycolysis of glucose of rat brain tissues by interrupting the conversion of glucose to glucose-6-phosphate, catalyzed by the enzyme hexokinase. Morphine, in the same concentrations, has no effect. Methadone injected intravenously into rats shows practically no age difference in toxicity, while morphine is especially toxic to weaning or aged animals.⁴³ Clinical reports^{10, 12} indicate that euphoria, which occurs in non-addicts following morphine, is rarely present, if at all, with methadone. Although physical dependence develops with methadone, its withdrawal symptoms are so slight that it can be used for the treatment of morphine addiction, resulting in a smooth, milder abstinence period.^{12, 44, 45} The rhesus monkey, like man, is easily susceptible to morphine, but it does not show tolerance or addiction to methadone, as proved by Woods, Wyngaarden, and Seever.^{36, 37}

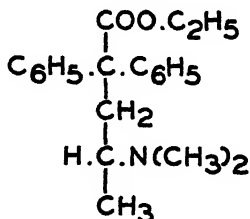
Optical Isomers. There is an asymmetrical C-atom in the molecule of methadone, and several papers on the difference in potency of the optical isomers have appeared.⁴⁶⁻⁵¹ Certain investigators believe that *l*-methadone* is quantitatively more potent than the *d*-form, while others claim that the *d*-isomer is entirely inactive. In our laboratory, it was found that, in rats, *l*-methadone is 7 times as active as the *d*-isomer.⁵² The difference becomes greater in larger animals, for the *l*-isomer is 25 times as active in dogs, and 50 times as active in man, as the *d*-form.

The intravenous toxicity in mice presents peculiarities.⁵² Numerically, there is no significant difference between the median lethal doses of *d*- and *l*-methadone, but the racemic mixture is definitely more toxic than either optical isomer. In those experiments, it was observed that *l*-methadone in lethal or sublethal doses produced protracted prostration; it caused delayed death. With the *d*-isomer, the animal promptly went

* *l*-Methadone merely indicates levo-rotation; it may not have *l*-configuration until it has been stereochemically proved and confirmed.

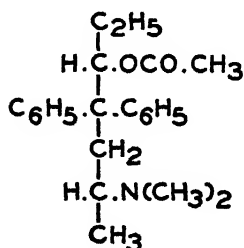
into convulsions and either died or recovered quickly. It is, therefore, not improbable that a synergism of toxicity exists between the two optical isomers. The racemic mixture is then capable of causing both slow and quick death.

Another pair of optically active isomers was compared for their analgesic potency. They are *l*- and *d*-ethyl-4-dimethylamino-2,2-diphenyl valerate hydrochlorides:



They were prepared from the same intermediates as used for the optical isomers of methadone. The interesting feature is that, in rats, by subcutaneous injection the *d*-isomer of the ethyl ester is 1/5 as active, and the *l*-form, 1/35 as active, as *dl*-methadone. Thus, the *d*-isomer is 7 times as potent as its *l*-antipode. There is practically no difference in their intravenous toxicity in mice, being 37.2 ± 2.5 and 37.3 ± 2.9 mg. per kg. for the *d*- and *l*-isomers, respectively.

A third pair of enantiomorphs, α -*l*- and α -*d*-3-acetoxy-4,4-diphenyl-6-dimethylaminoheptane hydrochlorides, was also investigated:



As shown in the structure, there are two asymmetrical C-atoms, and thus four possible optical isomers. Only one pair, designated as the α -pair, was available for our work. The *l*-antimer was made from *d*-methadone, and the *d*-antimer from *l*-methadone. In rats, by subcutaneous injection, the *d*-isomer is 5.4 times as active as the *l*-form, and twice as active as *dl*-methadone, analgesically. The *l*-isomer has a delayed onset, but a long duration, of action as compared with the *d*-isomer.

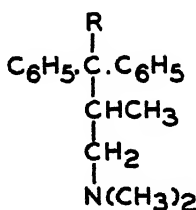
Derivatives of Methadone. The Germans prepared a good number of derivatives of methadone, several of which look interesting.¹ Workers in this country and in England have reported preliminary results of their

ture of the acetoxyl derivative (No. 27) is 30 per cent more active than methadone. The high potency of this compound has been previously reported by Sherrod, Kaiser, Santos-Martinez, and Pfeiffer.⁵⁵ The propionoxy analogue (No. 28) has the same activity as methadone in rats. The significance of the optical rotation of compounds No. 25 and No. 27 has been discussed above.

If a methyl group is attached to the β -C-atom, the resulting variants, listed in TABLE 7, are of more than casual interest. The ketone (No. 31)

TABLE 7

Compound No.	R	Activity in rats: % of methadone
30	COOH	0
31	CO.C ₂ H ₅	30
32	COO.C ₂ H ₅	17
33	OOC.CH ₃	17
	—CH.C ₂ H ₅	



has been studied and reported on by different workers under the name of isomethadone.^{47-55, 57} The results of our own tests indicate that it is about half as potent as methadone when injected subcutaneously in rats.

Bibliography

1. KLEIDERER, E. C., J. B. RICE, & V. CONQUEST. 1945. Pharmaceutical activities at the I. G. Farbenindustrie plant, Höchst-am-Main, Germany. Rept. 981: Office of the Publication Board, Dept. of Commerce, Washington, D. C.
2. Council on Pharmacy and Chemistry. 1947. Methadon, generic term for 6-dimethylamino-4,4-diphenyl-3-heptanone. J. A. M. A. 134: 1483.
3. SCOTT, C. C., & K. K. CHEN. 1946. The action of 1,1-diphenyl-1-(dimethylaminoisopropyl)-butanone-2, a potent analgesic agent. Fed. Proc. 5: 201.
1946. The action of 6-dimethylamino-4,4-diphenyl-3-heptanone, a potent analgesic agent. J. Pharm. & Exp. Therap. 87: 63.
4. HAFNER, F. 1929. Experimentelle Prüfung schmerzstillender Mittel. Deut. med. Wochenschr. 55: 731.
5. ANDREWS, H. L., & W. WORKMAN. 1941. Pain threshold measurements in the dog. J. Pharm. & Exp. Therap. 73: 99.
6. HARDY, J. D., H. G. WOLFF, & H. GOODALL. 1940. Studies on pain. A new method for measuring pain threshold: observations on spatial summation of pain. J. Clin. Invest. 19: 649.
7. SCOTT, C. C., K. G. KOHLSTAEDT, & K. K. CHEN. 1947. Comparison of the pharmacologic properties of some new analgesic substances. Anesth. & Analg. 26: 12.
8. SCOTT, W. W., H. M. LIVINGSTONE, J. J. JACOBY, & G. R. BROBERG. 1947. Early clinical experience with Dolophine (No. 10820). Anesth. & Analg. 26: 18.
9. KIRCHHOFF, A. C., & N. A. DAVID. 1947. Clinical trial of a new synthetic heptanone analgesic (Dolophine). Preliminary report. West. J. Surg. 55: 183.
1947. Clinical trial of 6-dimethylamino-4,4-diphenyl-3-heptanone (Dolophine), a synthetic analgesic. Fed. Proc. 6: 345.
1948. A clinical trial of amidone (Dolophine), a new synthetic analgesic. Anesth. & Analg. 27: 92.

10. TROXIL, E. B. 1947. The analgesic action of 1,1-diphenyl-1-(dimethylaminoisopropyl)-butanone-2 in man. *Fed. Proc.* 6: 378.
1948. Clinical evaluation of the analgesic methadon. *J. A. M. A.* 136: 920.
11. GENTLING, A. A., & J. S. LUNDY. 1947. A new analgesic AN-148: preliminary report. *Proc. Staff Meet. Mayo Clinic* 22: 249.
12. ISBELL, H., A. WIKLER, N. B. EDDY, J. L. WILSON, & C. F. MORAN. 1947. Tolerance and addiction liability of 6-dimethylamino-4,4-diphenyl-heptanone-3 (methadon). *J. A. M. A.* 135: 888.
13. BERTEL, N. A. 1948. Clinical trial of 10820, a new synthetic analgesic. *Dis. Nervous System* 9: 15.
14. POPKIN, R. J. 1948. Experiences with a new synthetic analgesic, amidone. Its action on ischemic pain of occlusive arterial diseases. *Am. Heart J.* 35: 793.
15. HENDERSON, F. G., E. R. ROBBINS, & K. K. CHEN. Action and toxicity of methadon in different animals. (To be published.)
16. ISBELL, H., A. WIKLER, A. J. EISENMAN, & K. FRANK. 1947. Effect of single doses of 10820 (4,4-diphenyl-6-dimethylamino-heptanone-3) on man. *Fed. Proc.* 6: 341.
17. ISBELL, H., A. J. EISENMAN, A. WIKLER, & K. FRANK. 1948. The effects of single doses of 6-dimethylamino-4,4-diphenyl-3-heptanone (amidone, methadon, or '10820') on human subjects. *J. Pharm. & Exp. Therap.* 92: 83.
18. SCOTT, C. C., K. G. KOHLSTAEDT, E. B. ROBBINS, & F. W. ISRAEL. 1947. Further observations on the pharmacology of Dolophine. *Fed. Proc.* 6: 370.
19. SCOTT, C. C., K. K. CHEN, K. G. KOHLSTAEDT, E. B. ROBBINS, & F. W. ISRAEL. 1947. Further observations on the pharmacology of 'Dolophine' (methadon, Lilly). *J. Pharm. & Exp. Therap.* 91: 147.
20. HEINIKAMP, W. J. R. 1922. The mechanism of the Straub biologic test for morphine. *J. Pharm. & Exp. Therap.* 20: 107.
21. WIKLER, A., K. FRANK, & A. J. EISENMAN. 1947. Effects of single doses of 10820 (4,4-diphenyl-6-dimethylamino-heptanone-3) on the nervous system of dogs and cats. *Fed. Proc.* 6: 384.
22. LEIMDORFER, A. 1948. An electroencephalographic analysis of the action of amidone, morphine and strychnine on the central nervous system. *Arch. Internat. Pharmacodyn. & Therap.* 76: 153.
23. EISLER, O., & O. SCHAUMANN. 1939. Dolantin, ein neuartiges Spasmolytikum und Analgetikum. *Deut. med. Wochenschr.* 65: 967.
24. SCHAUMANN, O. 1940. Ueber eine neue Klasse von Verbindungen mit spasmolytischer und zentral analgetischer Wirksamkeit unter besonderer Berücksichtigung des 1-Methyl-4-phenyl-piperidin-4-carbonsäure-äthylesters (Dolantin). *Arch. Exp. Path. & Pharmacol.* 196: 109.
25. KIRCHHOFF, A. C. & J. K. UCHIYAMA. 1947. Spasmolytic action of 6-dimethylamino-4,4-diphenyl-3-heptanone (Dolophine), a synthetic analgesic. *Fed. Proc.* 6: 345.
26. UCHIYAMA, J., A. C. KIRCHHOFF, & N. A. DAVID. 1947. Spasmolytic action of Dolophine (amidone). *Proc. Soc. Exp. Biol. & Med.* 66: 417.
27. GRUBER, C. M., E. R. HART, & C. M. GRUBER, JR. 1941. The pharmacology and toxicology of the ethyl ester of 1-methyl-4-phenyl-piperidine-4-carboxylic acid (Denierol). *J. Pharm. & Exp. Therap.* 73: 319.
28. POWELL, C. E., & K. K. CHEN. 1948. Action of dolophine on the animal uterus. *J. Am. Pharm. Assoc.* (In press.)
29. LEE, H. M. Unpublished data.
30. TRATNER, H. R. 1924. A method for recording contractions of the intact human ureter. *J. Urol.* 11: 444.
31. EVERETT, F. G. 1948. The local anesthetic properties of amidone (Dolophine). *Anesthesiology* 9: 115.
32. HAAG, H. B., J. K. FINNEGAN, & P. S. LARSON. 1947. Pharmacologic observations on 1,1-diphenyl-1-(dimethyl-aminoisopropyl)-butanone-2. *Fed. Proc.* 6: 334.
33. FINNEGAN, J. K., H. B. HAAG, P. S. LARSON, & M. L. DREYFUSS. 1948. Observations on the comparative pharmacologic actions of 6-dimethylamino-4,4-di-

- phenyl-3-heptanone (amidone) and morphine. *J. Pharm. & Exp. Therap.* **92**: 269.
34. PHATAK, N. M., J. MALONEY, & N. DAVID. 1948. Use of hyperglycemic response for estimating addiction potentialities of analgesic compounds. *Fed. Proc.* **7**: 249.
 35. CRONHEIM, G., & P. A. WARE. 1948. The determination and the urinary excretion of 6-dimethylamino-4,4-diphenyl-3-heptone hydrochloride (amidone). *J. Pharm. & Exp. Therap.* **92**: 98.
 36. WOODS, L. A., J. B. WYNGAARDEN, & M. H. SLEEVES. 1947. The addiction potentialities of 1,1-diphenyl-1-(dimethylaminoisopropyl)-butane-2 (amidone) in the monkey. *Fed. Proc.* **6**: 387.
 37. WOODS, L. A., J. B. WYNGAARDEN, & M. H. SLEEVES. 1947. Addiction potentialities of 1,1-diphenyl-1-(β -dimethylaminopropyl)-butanone-2 hydrochloride (amidone) in the monkey. *Proc. Soc. Exp. Biol. & Med.* **65**: 113.
 38. ISBELL, H., A. J. EISENMAN, A. WIKLER, M. DAINGERFIELD, & K. FRANK. 1947. Experimental addiction to 10820 (4,4-diphenyl-6-dimethylamino-heptanone-3) in man. *Fed. Proc.* **6**: 264.
 39. WIKLER, A., & K. FRANK. 1947. Tolerance and physical dependence in intact and chronic spinal dogs during addiction to 10820 (4,4-diphenyl-6-dimethylamino-heptanone-3). *Fed. Proc.* **6**: 384.
 40. SHIDEMAN, F. E., & H. T. JOHNSON. 1947. Acute vascular tolerance to morphine, Demerol, and 1,1-diphenyl-1-(dimethylamino-isopropyl) butanone-2 (amidone) in the dog. *Fed. Proc.* **6**: 371.
1948. Acute vascular tolerance to morphine, isonipecaine (Demerol), and methadon (amidone) in the dog. *J. Pharm. & Exp. Therap.* **92**: 414.
 41. KARR, N. W. 1947. Effects of 6-dimethylamino-4,4-diphenyl-3-heptanone (Dolophine) on intestinal motility. *Fed. Proc.* **6**: 343.
 42. GREIG, M. E. 1948. The inhibition of hexokinase by amidone (2-dimethylamino-4,4-diphenylheptanone-3-hydrochloride). *Arch. Biochem.* **17**: 129.
 43. HENDERSON, F. G., & K. K. CHEN. 1948. Effect of age upon the toxicity of methadon. *Proc. Soc. Exp. Biol. & Med.* (In press.)
 44. ISBELL, H., A. J. EISENMAN, A. WIKLER, M. DAINGERFIELD, & K. FRANK. 1947. Treatment of the morphine abstinence syndrome with 10820 (4,4-diphenyl-6-dimethylamino-heptanone-3). *Fed. Proc.* **6**: 340.
 45. EDDY, N. B. 1947. A new morphine-like analgesic. *J. Am. Pharm. Assoc.* **8**: 536.
 46. THORP, R. H., E. WALTON, & P. OFNER. 1947. Optical isomers of amidone, with a note on *iso*amidone. *Nature* **160**: 605.
 47. ISBELL, H., & A. J. EISENMAN. 1948. Physical dependence liability of drugs of the methadon series and of 6-methyldihydromorphine. *Fed. Proc.* **7**: 162.
 48. DENTON, J. E., O. H. STRAUS, W. E. WADDELL, & H. K. BECKER. 1948. A comparison of side actions and analgesic effects of morphine, amidone (4,4-diphenyl-6-dimethylamino-heptanone-3) and its isomers in man. *Fed. Proc.* **7**: 214.
 49. HOPPE, J. O., & L. C. MILLER. 1948. A comparison of the toxicity of methadon and some related compounds. *Fed. Proc.* **7**: 228.
 50. JENNEY, E. H., & C. C. PFEIFFER. 1948. Comparative analgesic and toxic effects of the optical isomers of methadon and isomethadon. *Fed. Proc.* **7**: 231.
 51. LUDUENA, F. P., L. C. MILLER, E. ANANENKO, & J. D. FRICK. 1948. Some pharmacological actions of isomers of methadon. *Fed. Proc.* **7**: 241.
 52. SCOTT, C. C., E. B. ROBBINS, & K. K. CHEN. 1948. Pharmacologic comparison of the optical isomers of 'Dolophine' (Methadon, Lilly). *J. Pharm. & Exp. Therap.* (In press.)
 53. SCOTT, C. C., E. B. ROBBINS, & K. K. CHEN. 1946. Comparison of some new analgesic compounds. *Science* **104**: 387.
 54. THORP, R. H., E. WALTON, & P. OFNER. 1947. Analgesic activity in compounds related to amidone. *Nature* **159**: 679.
 55. SHERROD, T., R. KAISER, J. SANTOS-MARTINEZ, & C. C. PFEIFFER. 1948. Methadon derivatives of pharmacological interest. *Fed. Proc.* **7**: 255.

56. KIRCHHOF, A. C. 1948. Further studies on synthetic analgesics. *Fed. Proc.* 7: 234.
57. EASTON, N. R., J. H. GARDNER, M. L. EVANICK, & J. R. STEVENS. 1948. Some isomers of amidone and related compounds. *J. Am. Chem. Soc.* 70: 76.
58. BLICKE, F. F., & J. KRAPCHO. 1948. Cyclic analogs of amidone. *Abst., 113th Meet., Am. Chem. Soc.*: 3K.
59. CHENEY, L. C., R. R. SMITH, & S. B. BINKLEY. 1948. Analgesics, ketimines and acylimines related to amidone. *Abst., 113th Meet., Am. Chem. Soc.*: 2K.
60. SPEETER, M. E., W. M. BYRD, & L. C. CHENEY. 1948. Analgesic carbinols and esters related to amidone. *Abst., 113th Meet., Am. Chem. Soc.*: 1K.
61. DODDS, E. C., W. LAWSON, & P. C. WILLIAMS. 1943. Morphine-like properties of diphenylethylamine and related compounds. *Nature* 151: 614.
1944. Morphine-like properties of diphenylethylamine and related compounds. *Proc. Roy. Soc.* B132: 119.

ANIMAL EXPERIMENTATION IN STUDYING ADDICTION TO THE NEWER SYNTHETIC ANALGESICS

By M. H. SEEVERS

Department of Pharmacology, University of Michigan, Ann Arbor, Michigan

THE term addiction, when applied to the syndrome resulting from the prolonged use of morphine and its congeners in man, by common usage involves three principal components, *psychic dependence*, *physical dependence*, and acquired *tolerance*. In view of the fact that the newer synthetic analgesics produce pharmacological effects, many of which are qualitatively similar to those of morphine and its derivatives, it is assumed, for purposes of this discussion, that morphine should stand as the drug of comparison, and that information regarding these three not easily dissociable qualities which constitute addiction to morphine must be obtained for each new agent of this type before it may be safely used clinically.

It is clear that clinical facilities for evaluating addiction liability are inadequate to keep pace with the rapid development of the newer synthetic analgesics. This being the case, and recognizing that the final evaluation of addiction liability of any compound proposed for clinical use must be determined in the human species, only that evidence which relates to the following question will be presented and discussed here: Does any infrahuman species exist which parallels normal man in its ability to develop "addiction" when chronically poisoned with morphine, its congeners, and the newer synthetic analgesics, and, if so, is this parallelism sufficiently striking that this species could be used as a reliable test object for the preliminary screening of new compounds? Since comparatively little work has been done in this field with the newer analgesic agents, a large share of the following discussion must center about morphine.

Psychic Dependence

At the outset, it should be clearly stated that the term "addiction," when applied to the condition resulting from chronic administration of morphine to infrahuman animal species, must have a more limited and specific connotation than when such a term is applied to the human addict. It is, perhaps, unfortunate that the term "psychic dependence" as it applies to morphine addiction in man has, in the broadest and most widely used sense, come to imply two differing but related factors: (a) The characteristic and inherent pre-addiction urge of certain individuals to seek "euphoria" as a manifestation of an existing neurosis or psychosis. (Pescor¹ found 96.2 per cent of 1,036 addicts at the Lexington Hos-

pital to be either psychopathic or suffering from some neurosis prior to addiction.) (b) The psychosomatic conditioning established by prolonged administration of the drug which involves anticipation not only of the thrill, but also the mental relief from impending or actual physical symptoms of withdrawal.

Whereas certain higher animal species become conditioned to accepting the drug without the usual resistance and may even learn to associate the administration of the drug with the relief of the symptoms of abstinence, no counterpart of the manifestation of pleasure or "euphoria" obviously derived by addicts from repetitious use of these compounds has been observed in any animal species thus far studied.

Spragg² has shown that the chimpanzee will make an objective choice of the syringe over food during withdrawal, and the author has observed behavior in the rhesus monkey which leads him to believe that this animal can make a similar association. Although these observations in animals are of considerable interest as they relate to item *b* above, it seems quite clear that it would be futile to attempt to transfer such observations into the clinic or use them as a basis for deductions or predictions regarding *a*, *i.e.*, those qualities in the drug which would induce "euphoria," satisfy the "craving" of the inebriate type of personality, or appeal to the confirmed addict already conditioned by his previous experiences with drugs of this type.

Both meperidine and methadone have been administered to monkeys for fairly long periods of time without any evidence of desire on the part of the animals, even during withdrawal of these drugs. Such information should not, however, lead to the inference that addicts would not obtain a pleasurable sensation from the use of these drugs, a fact which has already been clearly established for methadone in the clinic. *For practical purposes, then, clinically applicable information from animals relating to addiction liability is limited to studies of tolerance and physical dependence.*

Physical Dependence

Physical dependence to morphine can be established in numerous animal species³ including the mouse, rat, guinea pig, rabbit, cat, dog, monkey, and chimpanzee. The nature of the signs observed is dependent upon the physical characteristics of the species, and the phylogenetic development of the central nervous system, generalized "hyperirritability" being the only constant sign of abstinence in the lower species. In the smaller animals, wide individual variation in intensity of signs is noted, even after prolonged administration of large doses. Attempts have been made to use small animals for screening purposes. Barlow's method of using an increase in pre-injection irritability in the rat has been reported upon favorably.^{4, 5} A wide variation in individual response was noted, a factor which in our hands greatly limits its value as a screening method, although it has not been applied to a study of the newer

agents. Recently Phatak, Maloney, and David⁶ suggested the use of the hyperglycemic response of the rabbit for estimating addiction potentialities of analgesic compounds, and made observations on a group of methadone congeners. Unfortunately, the rabbit does not show satisfactory or reproducible abstinence phenomena during morphine withdrawal, and man does not obtain a hyperglycemic response to methadone. The author believes that any objective method, to be useful for comparative purposes, must involve signs which are demonstrable in man as well as in the animal under study. This statement would apply to changes in weight, temperature, blood sugar, variations in contractility of smooth muscle, circulatory alterations, or any other objective signs.

Whereas small animals may be used satisfactorily for studies of tolerance development, and probably for certain studies regarding mechanism of addiction, the author believes that only the dog and the monkey show signs of withdrawal which are sufficiently comparable to those observed in man to be reliable for screening and for the evaluation of new drugs. In view of the limited space available here, an attempt will be made only to compare these two species.

The Dog. Many observations have been made on this species, particularly with morphine.³ The abstinence syndrome is, in general, similar to that noted in man in so far as signs can be compared in view of the anatomical dissimilarities of the two species. Unfortunately, wide individual variation in response exists in different animals irrespective of dosage or duration of poisoning. Codeine⁷ does not induce physical dependence, although signs of abstinence have been noted with dihydromorphinone.⁸ Although tolerance studies have been made with heroin, no data have been published concerning physical dependence in this species.

Barlow⁹ gave 8 dogs 75 mg./kg. of meperidine orally once daily for ten months; 4 dogs 15 mg./kg. intramuscularly every 8 hours for 28 days; and 7 dogs 4 mg./kg. intramuscularly every two hours day and night for 5 days. Other than the acute effects of the drug noted following administration, which included salivation, muscle tremors, spasticity, anorexia and weight loss, and an intense dislike for the drug which increased as the experiment progressed, no untoward signs were observed and no signs of abstinence were detected on withdrawal.

Scott and his co-workers^{10, 11} did not note any signs of withdrawal in dogs administered methadone for several weeks. Wikler and Frank¹² gave 1 to 2 mg./kg. of methadone 4 times daily to dogs, increasing the dose gradually to 5 mg./kg., and continued the experiment for 10 weeks. Abrupt withdrawal revealed a marked abstinence syndrome characterized by restlessness, severe muscle tremors, fever, tachycardia, vomiting, hyperpnea, hydrophilia, and loss of weight, which appeared within 10 hours, reached a maximum in 24 hours, and subsided almost completely

in 48 hours. Similar results were obtained in chronic spinal dogs and one chronic decorticated dog.¹³ In summary, Wikler states, "Quantitatively, signs of abstinence from methadon were easier to produce (could be developed in 1 month as compared with 2 or 3 months with morphine), came on more rapidly (apparent in 9 hours compared with 22 hours with morphine) and were more severe."

The Monkey. The abstinence syndrome in the rhesus monkey is objectively quite similar to that noted in man.¹⁴ Whereas individual variations exist in this species, and the observer must familiarize himself with the personality characteristics of each animal prior to addiction in order to make a satisfactory evaluation of abstinence phenomena (as is also the case with the dog), the picture is more uniform and more constantly reproducible than with the dog or any animal other than the chimpanzee. As in man, the monkey can exert wilful control over signs of abstinence. This may be related, in part, to fear of the attendant, whose entrance into the room produces changes in behavior which mask the signs of abstinence. When the animal is observed during abstinence through a sound-proof, "one-way vision" glass window, the true picture of abstinence is revealed. This technique, recently adopted in this laboratory, has greatly enhanced the reliability of our observations.

Heroin and dihydromorphinone produce signs of abstinence comparable to morphine, whereas codeine produces only minimal signs in this species.¹⁵

The only observations on meperidine known to the author to have been made in this species are those of Barlow.⁹ He administered meperidine to 14 rhesus monkeys in dosage of 14 mg./kg. orally once daily for 10 months; to 5 monkeys 15 mg./kg. intramuscularly every 8 hours for 28 days; and to 7 monkeys 4 mg./kg. intramuscularly every 2 hours for 5 days. Whereas these doses produced marked acute effects, no signs of abstinence were noted on withdrawal.

Woods, Wyngarden, and Seevers¹⁶ administered methadone in an initial dose of 5 mg./kg. once daily subcutaneously to monkeys. This dose was increased in 24 to 26 days to the maximum tolerated acute dose (11 to 13 mg./kg.), and then continued at this level for 75 to 96 days. No signs of abstinence were detected on withdrawal, although controls on morphine demonstrated the characteristic picture of animal addiction to this drug.

Cochin, Gruhzit, Woods, and Seevers¹⁷ repeated these experiments recently, but administered the drug 3 times daily in a dosage of 7 mg./kg. (total daily dose of 21 mg./kg.) for four and one-half months. This represented the maximum tolerated dose at this interval of administration. Controls with morphine were given 50 mg./kg. thrice daily. Signs of abstinence from methadone, if they actually existed at all, were minimal. Slight pilomotor activity was noted at 36 hours. The animals were more active at 36 hours, but this appeared to be merely the release from metha-

agents. Recently Phatak, Maloney, and David⁶ suggested the use of the hyperglycemic response of the rabbit for estimating addiction potentialities of analgesic compounds, and made observations on a group of methadone congeners. Unfortunately, the rabbit does not show satisfactory or reproducible abstinence phenomena during morphine withdrawal, and man does not obtain a hyperglycemic response to methadone. The author believes that any objective method, to be useful for comparative purposes, must involve signs which are demonstrable in man as well as in the animal under study. This statement would apply to changes in weight, temperature, blood sugar, variations in contractility of smooth muscle, circulatory alterations, or any other objective signs.

Whereas small animals may be used satisfactorily for studies of tolerance development, and probably for certain studies regarding mechanism of addiction, the author believes that only the dog and the monkey show signs of withdrawal which are sufficiently comparable to those observed in man to be reliable for screening and for the evaluation of new drugs. In view of the limited space available here, an attempt will be made only to compare these two species.

The Dog. Many observations have been made on this species, particularly with morphine.³ The abstinence syndrome is, in general, similar to that noted in man in so far as signs can be compared in view of the anatomical dissimilarities of the two species. Unfortunately, wide individual variation in response exists in different animals irrespective of dosage or duration of poisoning. Codeine⁷ does not induce physical dependence, although signs of abstinence have been noted with dihydromorphinone.⁸ Although tolerance studies have been made with heroin, no data have been published concerning physical dependence in this species.

Barlow⁹ gave 8 dogs 75 mg./kg. of meperidine orally once daily for ten months; 4 dogs 15 mg./kg. intramuscularly every 8 hours for 28 days; and 7 dogs 4 mg./kg. intramuscularly every two hours day and night for 5 days. Other than the acute effects of the drug noted following administration, which included salivation, muscle tremors, spasticity, anorexia and weight loss, and an intense dislike for the drug which increased as the experiment progressed, no untoward signs were observed and no signs of abstinence were detected on withdrawal.

Scott and his co-workers^{10, 11} did not note any signs of withdrawal in dogs administered methadone for several weeks. Wikler and Frank¹² gave 1 to 2 mg./kg. of methadone 4 times daily to dogs, increasing the dose gradually to 5 mg./kg., and continued the experiment for 10 weeks. Abrupt withdrawal revealed a marked abstinence syndrome characterized by restlessness, severe muscle tremors, fever, tachycardia, vomiting, hyperpnea, hydrophilia, and loss of weight, which appeared within 10 hours, reached a maximum in 24 hours, and subsided almost completely

in 48 hours. Similar results were obtained in chronic spinal dogs and one chronic decorticated dog.¹³ In summary, Wikler states, "Quantitatively, signs of abstinence from methadon were easier to produce (could be developed in 1 month as compared with 2 or 3 months with morphine), came on more rapidly (apparent in 9 hours compared with 22 hours with morphine) and were more severe."

The Monkey. The abstinence syndrome in the rhesus monkey is objectively quite similar to that noted in man.¹⁴ Whereas individual variations exist in this species, and the observer must familiarize himself with the personality characteristics of each animal prior to addiction in order to make a satisfactory evaluation of abstinence phenomena (as is also the case with the dog), the picture is more uniform and more constantly reproducible than with the dog or any animal other than the chimpanzee. As in man, the monkey can exert wilful control over signs of abstinence. This may be related, in part, to fear of the attendant, whose entrance into the room produces changes in behavior which mask the signs of abstinence. When the animal is observed during abstinence through a sound-proof, "one-way vision" glass window, the true picture of abstinence is revealed. This technique, recently adopted in this laboratory, has greatly enhanced the reliability of our observations.

Heroin and dihydromorphinone produce signs of abstinence comparable to morphine, whereas codeine produces only minimal signs in this species.¹⁵

The only observations on meperidine known to the author to have been made in this species are those of Barlow.⁹ He administered meperidine to 14 rhesus monkeys in dosage of 14 mg./kg. orally once daily for 10 months; to 5 monkeys 15 mg./kg. intramuscularly every 8 hours for 28 days; and to 7 monkeys 4 mg./kg. intramuscularly every 2 hours for 5 days. Whereas these doses produced marked acute effects, no signs of abstinence were noted on withdrawal.

Woods, Wyngarden, and Seevers¹⁶ administered methadone in an initial dose of 5 mg./kg. once daily subcutaneously to monkeys. This dose was increased in 24 to 26 days to the maximum tolerated acute dose (11 to 13 mg./kg.), and then continued at this level for 75 to 96 days. No signs of abstinence were detected on withdrawal, although controls on morphine demonstrated the characteristic picture of animal addiction to this drug.

Cochin, Gruhzit, Woods, and Seevers¹⁷ repeated these experiments recently, but administered the drug 3 times daily in a dosage of 7 mg./kg. (total daily dose of 21 mg./kg.) for four and one-half months. This represented the maximum tolerated dose at this interval of administration. Controls with morphine were given 50 mg./kg. thrice daily. Signs of abstinence from methadone, if they actually existed at all, were minimal. Slight pilomotor activity was noted at 36 hours. The animals were more active at 36 hours, but this appeared to be merely the release from metha-

done depression and a return to normal excitability rather than a state of hyperexcitability. Anorexia was not noted as was the case with morphine withdrawal, and no significant weight change occurred. In contrast, the controls on morphine showed the most marked signs of withdrawal ever noted by the author. This was probably related to the rate of administration, since virtually all previous experiments in the monkey have been made with single daily dosage. One animal, after a period in which he exhibited all the classical signs of withdrawal, lapsed into a state of profound exhaustion, refused food and water, became intensely dyspneic, and, after some struggling incident to the removal from the cage after 76 hours for movie recording, became intensely cyanotic and died within 15 minutes. This death was evidently of circulatory type, the heart being acutely dilated at autopsy. This animal was otherwise in good physical state, had not lost weight, and no other pathology was revealed at autopsy.

Tolerance

Whereas acquired tolerance is always demonstrable and a part of the picture of morphine addiction, its exact relationship to physical dependence and psychic dependence is not clearly defined. That the acquisition of tolerance is not necessary to the establishment of psychic dependence to a drug is clearly shown in the case of cocaine, since the latter may be clearly established without any evidence of tolerance development. Conversely, marked tolerance may be acquired to other compounds, such as the organic nitrites, without associated psychic or physical dependence.

It is evident that various body mechanisms and tissues differ qualitatively and quantitatively with respect to tolerance development.

Blood vessels become rapidly tolerant to morphine, as was first shown by Schmidt and Livingstone.¹⁸ Recently, Shideman and Johnson¹⁹ have shown that some vascular tolerance can be developed to meperidine and methadone, and they have compared it with morphine in the dog. In a general way, this "acute tolerance" parallels the development of "chronic tolerance" to the sedative or narcotic action of these three drugs, both appearing rapidly with morphine, more slowly with methadone, and poorly and incompletely with meperidine. Whether this method can be used for screening to predict tolerance development awaits more extended studies with a greater number of new compounds.

Contrary to the situation with respect to physical dependence, small animals may be used satisfactorily for estimation of tolerance to the analgesic and general sedative action of these drugs.

Tolerance to the *analgesic action* of morphine and its derivatives is readily established in most species of animals. The author is not aware of carefully controlled studies on animals with meperidine regarding tolerance to its analgesic effects. Tolerance develops in former morphine

addicts,²⁰ but this is not a prominent feature of its clinical use in non-addicts.²¹ Tolerance to the analgesic action of methadone has been developed in the mouse,²² the rat,^{11, 23} the dog,¹² and in man.²²

Tolerance to the *general depressant action* is readily established for morphine and its derivatives in all species.³ Some tolerance is developed to this property for meperidine in rats and dogs although it is not marked.⁹ Tolerance to the sedative action of methadone is readily developed in the rat,¹⁰ in the dog,¹² slowly in the monkey¹⁶ and man,²² especially to larger dosages.

Acquired tolerance does not raise the *lethal dose* of morphine derivatives in any animal lower than the monkey, since death is due to convulsions rather than respiratory depression and tolerance is not developed to this stimulant action. This is probably true also of meperidine and methadone, since these drugs also produce death of a convulsant type in smaller animals. Since some monkeys and all men die of primary respiratory depression from morphine, it is easy to establish tolerance to this respiratory depressant effect of morphine, heroin and Dilaudid (not codeine), and consequently to what would otherwise be a lethal dose. The monkey (and presumably man) always succumbs to respiratory failure with lethal doses of methadone.¹⁶ Very little, if any, tolerance is developed to this respiratory depressant effect and a small increment in dosage above the maximum tolerated single dose, 12 to 15 mg./kg., will result in respiratory failure. We¹⁷ have obtained some suggestive evidence, which will require further study and confirmation, that respiratory tolerance to morphine confers crossed tolerance to methadone. By mistake, morphine-addicted animals were given 25 mg./kg. of methadone, a dose which we believe to be certainly lethal. Whereas these monkeys were treated after about thirty minutes with caffeine, it is believed that they would not have succumbed without such treatment. It is interesting, in this regard, that all of the evidence which indicates a low-grade respiratory tolerance to methadone in man has been obtained on addicts with a highly developed tolerance to morphine.

The author believes that predictions regarding addiction liability cannot be made at this time on the basis of tolerance studies alone, although it seems clear that the degree of addiction liability with morphine, meperidine, and methadone parallels roughly the degree of tolerance development, especially to the narcotic action of large doses of these drugs.

Summary

In considering the problem of predicting addiction liability in man from animal experiments, the conclusion is inevitable that the only reliable criterion upon which to base a decision is the establishment or lack of establishment of *physical dependence*. Up to the present time, no evidence is available which will refute the view long held by the author that the monkey is the best animal species available *practically* for the

reliable prediction of addiction liability in the "normal" human subject. This statement is based upon the following facts which hold for the normal monkey and normal man.

(a) The signs—and, judging from the actions of the monkeys, the symptoms—of abstinence from morphine, their time of appearance, their intensity, and duration, are for practical purposes identical in both species.

(b) The intensity of abstinence phenomena with the morphine derivatives, heroin, Dilaudid, and codeine are quantitatively similar.

(c) Signs of abstinence from meperidine are minimal.

(d) Complete and adequate substitution is readily established from morphine to its derivatives and to methadone in both the monkey and human addict.

(e) Tolerance development to these various drugs is qualitatively and quantitatively similar in both species.

(f) Physical dependence to methadone, if present at all, is slight and slowly developed in individuals not previously addicted to morphine.

It is hardly to be expected that studies of addiction liability in the "normal" monkey could be compared directly with studies made in human individuals who have been previously addicted to morphine. Such individuals not only represent a selected group who are psychiatrically inferior, but they have in addition been conditioned or "sensitized" to the whole experience of addiction, including their own selective interpretation of "euphoria."

The question is still pertinent, and as yet unanswered, whether prolonged addiction to morphine leaves a permanent physical as well as a mental residue even after apparent "cure." Thus, we do not know whether it is safe to make predictions, on the basis of addiction and substitution studies in actual or in "cured" morphine addicts, regarding the addiction liability, or even the "euphoria"-inducing qualities of new and chemically different compounds, if such predictions are to apply to psychiatrically "normal" individuals. It is probable that no studies on animals will ever solve the problem of whether a chemical compound is potentially addicting for "constitutional addicts." The published evidence seems to support the view that, excluding this group of individuals, normal human subjects, like normal monkeys (see TABLE 1), develop very little psychic or physical dependence to those synthetic analgesics, particularly meperidine and methadone, the only compounds which have thus far been even partially evaluated.

If such are the actual facts, it then seems evident that our problem in the evaluation of new compounds should be much broader than the mere search for a substance which will give perfect analgesia to the "normal" human subject without the development of the syndrome which we term addiction. We should emancipate our thinking from the concept that the capacity of a chemical substance to produce "euphoria" is inevitably an undesirable quality and make a positive and specific

TABLE 1
PHYSICAL DEPENDENCE TO ANALGESIC DRUGS

<i>Drug</i>	<i>Dog</i>	<i>Monkey</i>	<i>"Normal" human subjects</i>	<i>Human addicts</i>
Morphine	+++ (3)	++++ (3)	++++ (3)	++++ (3)
Heroin	{ not reported }	++++ (15)	++++ (3)	++++ (3)
Dihydro- morphinone	++ (8)	++++ (15)	++++ (3)	++++ (3)
Codeine	none (7)	+(15) *	? (3)	+(3)
Meperidine	none (9)	none (9) *	? (21)	+(21)
Methadone	{ +++++ (10-24 hours) (12) }	+? (16) *	{ not reported (22) }	{ ++ (fifth to ninth day) (22) }

* With maximum tolerated doses.

search for compounds which, devoid of the ability to induce *physical dependence*, will satisfy that large group of constitutionally inferior individuals now classified as criminals because they yield to an inherent psychic drive not under their control.

Bibliography

1. FESCOR, M. J. 1939. Public Health Rep. Suppl. 155.
2. SPRAGG, S. D. S. 1940. Comp. Psychol. Monogr. 15: 132.
3. KRUEGER, H., N. B. EDDY, & M. SUMWALT. 1941. Pub. Health Rep. Suppl. 165.
4. HIMMELSBACH, C. K., G. H. GERLACH, & E. J. STANTON. 1935. J. Pharm. & Exp. Therap. 53: 179-188.
5. EDDY, N. B., & C. K. HIMMELSBACH. 1936. Pub. Health Rep. Suppl. 118.
6. PHATAK, N. M., J. MALONEY, & N. DAVID. 1948. Fed. Proc. 7: 247.
7. TATUM, A. L., M. H. SEEVERS, & K. COLLINS. 1929. J. Pharm. & Exp. Therap. 36: 447.
8. EDDY, N. B., & J. S. REID. 1934. J. Pharm. & Exp. Therap. 52: 468-493.
9. BARLOW, O. W. Personal communication of unpublished experiments.
10. SCOTT, C. C., & K. K. CHEN. 1946. J. Pharm. & Exp. Therap. 87: 63.
11. SCOTT, C. C., K. S. KOHLSTAEDT, E. B. ROBBINS, & F. W. ISRAEL. 1947. Fed. Proc. 6: 370.
12. WIKLER, A., & K. FRANK. 1947. Fed. Proc. 6: 384.
13. WIKLER, A. 1945. Fed. Proc. 4: 141; 1946. *Ibid.* 5: 215.
14. SIEVERS, M. H. 1936. J. Pharm. & Exp. Therap. 56: 147.
15. SEEVERS, M. H. 1936. J. Pharm. & Exp. Therap. 56: 157.
16. WOODS, L. A., J. B. WYNGAARDEN, & M. H. SEEVERS. 1947. Proc. Soc. Exp. Biol. & Med. 65: 113.
17. COCHIN, J., C. C. GRUHZIT, L. A. WOODS, & M. H. SEEVERS. (To be published.)
18. SCHMIDT, C. F., & A. E. LIVINGSTON. 1933. J. Pharm. & Exp. Therap. 47: 443.
19. SHIDEMAN, F. E., & H. T. JOHNSON. 1948. J. Pharm. & Exp. Therap. 92: 414.
20. ANDREWS, H. L. 1942. J. Pharm. & Exp. Therap. 75: 333.
21. BITTERMAN, R. C., & C. K. HIMMELSBACH. 1943. J. A. M. A. 122: 222.
22. ISBI LI, H., A. WIKLER, N. B. EDDY, J. L. WILSON, & C. F. MORAN. 1947. J. A. M. A. 135: 888.
23. LEWIS, J. Personal communication.

Discussion of the Paper

DR. ABRAHAM WIKLER (*U. S. Public Health Service Hospital, Lexington, Kentucky*):

Although Dr. Seevers's paper deals primarily with animal experimentation, I should like to comment first on his conclusions as they apply to man. Dr. Seevers states that conclusions derived from studies on "psychiatrically inferior" individuals previously addicted to morphine cannot be expected to apply to "normal" persons not previously addicted. In this I should like to concur, if we consider the addiction problem only in its relation to "normal" persons. However, the normal well-adjusted, mature person is not likely to become a habitual drug addict. In our society, with legal and social attitudes such as they are, drug addiction becomes a public health problem chiefly in relation to those individuals whose emotional need for morphine, or drugs like it, is so strong that it overbalances the personality defenses (*e.g.*, "super-ego" structure) against addiction. Prominent in this group are the extremely infantile, narcissistic individuals who constitute the bulk of those post-addicts who volunteer for addiction studies. From a public health standpoint, it appears quite appropriate to study drug addiction in such individuals.

In commenting on the animal work which Dr. Seevers has discussed, I should like to limit my comments to some of the neurophysiological aspects, with particular reference to morphine and methadone, since my interest in addiction problems has been largely in this area. First, the terms "psychic" and "physical" dependence have acquired various connotations and should be defined more precisely. Thus, a distinction between "psychic" and "physical" cannot be made on the basis of objectivity or subjectivity of the symptom or sign. Tachycardia, vomiting, even fever, may be "psychic" in origin, while purposive behavior may be "physical" in origin, as has been demonstrated so strikingly by Richter's work on self-regulatory behavior in animals. Defined with reference to genesis, the distinction between "psychic" and "physical" can be made on the basis that the former is related to factors with symbolic significance varying in complexity, whereas the latter is related to factors with little or no symbolic significance. Viewed in these terms, it becomes apparent that theoretically, at least, it is not possible to distinguish between those aspects of addiction which are "psychic" in origin and those which are "physical" in intact man, or even in intact animals. Clinical experience has shown that the morphine abstinence picture is definitely related to the personality of the addict. Likewise, in our investigations, the effects of morphine on adaptive behavior in intact dogs have been shown to depend on the "personality" of the animal. Perhaps this accounts for the great variability in the abstinence picture from dog to dog, since this species exhibits a wide range of personality types.

True "physical dependence" in the sense described can, however, be studied in animal preparations in which a portion of the central nervous

system is isolated from the remainder, as in chronic decorticated and chronic spinal dogs. In such preparations, the capacity for symbolization is markedly impaired, and the reactions to various stimuli are stereotyped and predictable with accuracy. Thus, in the decorticated animals, circling activity, irritability, temperature, pulse, respirations, and tooth pain-reaction threshold can be measured accurately. In the chronic spinal preparations, the knee jerk, ipsilateral extensor thrust, ipsilateral flexor reflex, and crossed extensor reflex, as well as spontaneous activity of the hindlimbs can be recorded reliably. During addiction to morphine or methadone, tolerance is manifested to the action of these drugs on certain of the reactions mentioned, and during withdrawal, striking hyperactivity and other changes occur with great regularity and predictability in both types of preparations. The preparation and preservation of the chronic decorticated animals is a difficult procedure. However, chronic spinal dogs can be prepared in a one-stage operation. They require close attention during the first postoperative month, but after 4 to 6 weeks these preparations survive for years with a minimum of care, chiefly directed to keeping the hindlimbs and perineal regions dry. The reflexes attain a stable level in this period of time, and can be recorded very simply with an "isotonic" apparatus utilizing "natural" stimuli eliciting a maximal response. With this technique, the changes which occur spontaneously over a period of as long as 6 months are negligible compared to the effects of drugs during addiction studies.

Although the study of "physical" dependence can be facilitated by such techniques, the problem of species differences still remains. Whether or not the techniques described can be applied to monkeys will be determined by future investigations. That it would be of advantage to make such studies is apparent from Dr. Seevers's excellent review of the problem.

METHODS AND RESULTS OF STUDYING EXPERIMENTAL HUMAN ADDICTION TO THE NEWER SYNTHETIC ANALGESICS

By HARRIS ISBELL

Research Division, U. S. Public Health Service Hospital, Lexington, Kentucky

IN assessing the addiction liability of new drugs, attention must be paid to all three of the qualities which Himmelsbach and Small¹ have described as characteristic of addiction to the opiate drugs. These qualities are: tolerance, physical dependence, and habituation (emotional or psychic dependence). Tolerance is defined as a diminishing effect on repeated administration of the same dose of a drug. Physical dependence refers to an altered bodily state brought about by repeated administration of a drug over a long period of time, which necessitates continued use of the drug in order to prevent the appearance of a characteristic illness termed an abstinence syndrome. Habituation refers to emotional and psychological dependence on the drug—the substitution of the use of the drug for other methods of adaptive behavior. Although all observers are agreed that tolerance is less important than physical and emotional dependence, there is considerable disagreement concerning the relative importance of emotional and physical dependence. Pharmacologists are likely to hold the opinion that physical dependence is the only distinguishing characteristic of an addicting drug. Psychiatrists are likely to believe that emotional dependence is all-important in addiction. A position midway between these two extremes is probably best. Physical dependence and emotional dependence should be given equal weight in assessing addiction liability.

The final tests of the addiction liability of any new drug must be carried out on human subjects because animal experiments yield no information about the habituation liability of any drug, and because various species differ markedly in their susceptibility to the development of physical dependence on various drugs.

The addiction liability of a new drug is usually considered from two points of view: (1) What is the danger of addiction under conditions of legitimate medical use? (2) What is the danger that persons with susceptible personalities will illegally abuse the drug and so become addicted? The second question is much more important, because addiction seldom results from the legitimate medical use of a drug—even from the medical use of morphine. Every year, large numbers of people in the United States receive morphine for short periods of time and very few become addicted. Less than 5 per cent of patients admitted to the U. S. Public Health Service Hospital at Lexington, Kentucky, became addicted because drugs had been prescribed for them by physicians for the relief of

pain. The danger of addiction, under conditions of medical usage, is great only when physicians mistakenly believe that a drug is not addicting and are careless in its use. In the United States, addiction is chiefly a matter of abuse and is spread by persons already addicted. For this reason, tests of addiction liability at the U. S. Public Health Service Hospital are designed to simulate addiction under conditions of abuse. The assessment of addiction liability under conditions of medical use is usually carried out in other institutions.

Subjects for Tests of Addiction Liability. The men used as subjects for addiction-liability tests are selected from a group of men who have volunteered to undergo the experiments. These patients are all morphine addicts and are serving Federal sentences for violation of the Harrison Narcotic Act. All the experiments are explained to the patients in detail before they are begun, and participation is purely voluntary. Care is taken to choose men with poor prognoses for permanent cure of their addiction. Such patients have usually served a number of sentences for violation of the narcotic laws and have relapsed to the use of morphine after each discharge. No man can be used in an experiment unless he has six months or more of his sentence to serve after termination of the experiment in order that there will be time for the patient to recover from withdrawal of the drug. Objections are sometimes raised to the use of such subjects in testing addiction liability, on the grounds that former morphine addicts develop both physical and emotional dependence on a drug much more readily than do non-addicts. The answer to this objection is, of course, a practical one. Morphine addicts are the only persons who can ethically be used for addiction-liability experiments. Moreover, since we are testing addiction liability under conditions of abuse, the use of such "addiction-prone" subjects is an advantage, not a disadvantage. If experienced morphine addicts like the effects produced by a drug, that drug is dangerous, because the veteran addict is not only skilled in the methods of obtaining drugs illegally but also has few scruples about introducing non-addicts to the use of the drug.

Methods of Testing Addiction Liability. The methods used in testing addiction liability of new drugs are essentially those developed by Himmelsbach and his collaborators.¹⁻⁵ Four methods are available: administration of single doses of the drug under test to former morphine addicts for the detection of euphoria; the determination of the effect of single doses on the intensity of abstinence from morphine; substitution of the new drug for morphine in cases strongly addicted to morphine; and direct addiction.

Detection of Euphoria. Since most persons begin the use of drugs and become addicted because the drugs produce effects which they regard as pleasurable, the detection of euphoria is a very important procedure in evaluating addiction liability. The method used is simple: Single doses

of the drug under test are administered to former morphine addicts, and the subjects are unobtrusively watched for a period of 6 hours or more by specially trained observers. For our purposes, euphoria is defined as a series of effects similar to those produced by morphine. These effects are: increased talkativeness, boasting, greater ease in the experimental situation, expression of satisfaction with the effects of the drug, requests for increased doses of the drug, increased motor activity, and, with larger doses, slurring of speech, motor ataxia, and evidence of marked sedation. As many experiments are done as are necessary to reach a clear-cut conclusion. The observations are controlled by administering 30 mg. of morphine to the same subjects on other occasions. Initially, small subcutaneous doses of the drug under test are used, and if no untoward toxic effects are observed, the dosage is increased progressively in subsequent experiments until evidence of euphoria, roughly equivalent to that produced by 30 mg. of morphine, is detected, or, if no evidence of euphoria is detected, the dosage is elevated until further increases would be regarded as dangerous. If euphoria is detected, blind experiments are arranged in which neither the subject nor the observer are aware whether the drug given was morphine or the compound under test. Finally, various doses of the drug are administered intravenously. If a psychologist is available, administration of the Rorschach test and other projective tests before and after the new drug will yield valuable information about the euphoric qualities of the drug.

Relief of Abstinence from Morphine. Patients already strongly addicted to morphine are used in these experiments. Following a short preliminary test period of withdrawal to establish the presence of physical dependence, the patients are stabilized on the least amount of morphine which will prevent the appearance of signs of abstinence and are maintained on this level for 7 to 10 days. The morphine is then abruptly withdrawn and, beginning at the 24th hour of abstinence, hourly observations for the intensity of abstinence are made according to the hourly point score system of Himmelsbach³ (TABLE 1). In this system, arbitrary numerical values have been assigned to the various signs of abstinence and arbitrary limits set on certain signs—1 point for lacrimation; 3 points for mydriasis; 1 point for each increase in the respiratory rate, with a limit of 10, etc. Men with strong physical dependence on morphine usually score 20 to 30 points from the 24th to the 48th hours of abstinence. A dose of the drug under test is administered at the 28th to the 32nd hour, and the observations are continued. If there has been little effect, or no effect, after 4 hours, a second, larger dose may be given. If the drug lowers the intensity of abstinence, the observations are continued until the intensity of the withdrawal illness returns to the pre-dose level. If the drug does relieve abstinence, it is fairly likely that it will itself produce physical dependence.

Substitution for Morphine. Patients already addicted to morphine may

TABLE 1

METHODS AND RESULTS OF STUDYING EXPERIMENTAL
HUMAN ADDICTION TO THE NEWER SYNTHETIC ANALGESICS*Point system for measuring abstinence syndrome intensity by the day (D) or by the hour (H)*

<i>Signs</i>	<i>(D) by day</i>		<i>(H) by hour</i>	
	<i>Points</i>	<i>Limit</i>	<i>Points</i>	<i>Limit</i>
Yawning	1	1	1	1
Lacrimation	1	1	1	1
Rhinorrhea	1	1	1	1
Perspiration	1	1	1	1
Mydriasis	3	3	3	3
Tremor	3	3	3	3
Gooseflesh	3	3	3	3
Anorexia (40 per cent decrease in caloric intake)	3	3		
Restlessness	5	5	5	5
Emesis (each spell)	5		5	5
Fever (for each 0.1° C. rise over mean addiction level)	1		1	10
Hyperpnoea (for each resp./min. rise over mean addiction level)	1		1	10
Rise in A.M. Systolic B.P. (for each 2 mm. Hg over mean addiction level)	1	15	1	10
Weight loss (A.M.) (for each lb. from last day of addiction)	1			

Total abstinence syndrome intensity per day or per hour is the sum of the points scored in the (D) or (H) columns respectively, with due attention to the limits.

be used in these tests. Frequently, the same patients who serve as subjects for experiments on the effects of single doses on abstinence from morphine are also used as subjects for substitution studies. The presence of physical dependence on morphine is always proved in subjects accepted for this type of study by subjecting them to a 24- to 36-hour test period of withdrawal. If the patients show moderate to marked signs of abstinence during this period, they are returned to morphine, and the least dosage of morphine which will just prevent the appearance of signs of abstinence (the stabilization dosage) is determined by alternately raising or lowering the amount of morphine given. Once the stabilization dose is determined, the patients are maintained on that amount of drug for at least 7 days. The drug under test is then abruptly substituted for the morphine. If possible, the patients are kept unaware of the change. The dosage, and the interval of administration of the drug under test, are selected on the basis of the pharmacological data on the toxicity, relative potency, and length of action of the drug. If no signs of abstinence appear after the substitution has been effected, the dosage and interval of administration of the new drug are adjusted so as to determine the minimum amount of the drug which will prevent the appearance of signs of abstinence. Frequently, the drug under test will suppress signs of abstinence only partially, regardless of the dose used. Under such circumstances, mild signs of abstinence appear in the first few days of substitution, only to subside later. Occasionally, drugs are tested which will not support physical dependence at all. Such drugs have always, in

past experiments, been poor analgesics. After 7 to 14 days' substitution, the drug is abruptly and completely withdrawn. Observations for signs of abstinence are made every 2 hours in the first 48 hours of withdrawal, and 3 times daily thereafter. Patients are observed for 7 to 14 days after withdrawal, depending on the speed of onset and the rate of recovery from abstinence. The Himmelsbach² daily point scoring system is used in evaluating the intensity of abstinence after substitution and after withdrawal of the drug under test (TABLE 1). This system is identical with the hourly point scoring system, except that depression of caloric intake is scored and the limits set on certain signs are different. If a drug supports physical dependence on morphine, and if signs of abstinence are detected following withdrawal after substitution for morphine, the drug is regarded as having physical dependence liability. The Himmelsbach daily point score serves as an approximate measure of the physical dependence liability of the new drug as compared to morphine.

Direct Addiction. Direct addiction is the best method of determining the addiction liability of a new drug. It yields information about the development of tolerance to various actions of the new drug. The development of habituation can be followed, and the physical dependence liability of the drug is determined directly instead of by inference, as in the method using the effects of single doses on abstinence, or the substitution technique. The great disadvantage of the direct addiction method is the inordinate amount of time and labor involved. The studies must be carried on for periods of 3 to 7 months and require the full-time services of 5 attendants and 2 to 3 technicians.

Former morphine addicts who have been abstinent for 3 months or more are selected as study subjects from groups of volunteers. These men are studied intensively for 1 to 2 weeks before they are given any drugs. Physical and psychiatric examinations, various laboratory examinations, psychological tests, electroencephalograms, electrocardiograms, basal metabolic rates, observations of rectal temperature, pulse and respiratory rate, blood pressure, caloric intake, and hours of sleep are obtained in this period. The effects of single doses of the drug under test on the pain threshold, the electroencephalogram, etc., are also determined during this period.

Once the control observations have been obtained, administration of the drug is begun. At first, the dosages used are equal to those which would be used in clinical practice. The intervals of administration are based on the pharmacological data on the length of actions of the drug. Most commonly, drugs are administered every 6 hours. If, after a few days, no untoward toxic effects appear, the dosage is raised. Thereafter, the dosage is raised just as often as the patient's tolerance and the toxic effects of the drug will permit. Doses which pharmacologists regard as astronomical are usually reached. The elevation in dosage is eventually stopped, and the patients are maintained on the same dosage level for 2

weeks to 2 months prior to withdrawal so that they will be as completely tolerant as possible when the drug is withdrawn.

The observations made during the control period are repeated at intervals during the addiction period. Daily notes are written on the general behavior of the patients.

Detection of Tolerance. Tolerance to the sedative effects of the drug is easy to detect clinically. Checking the hours of sleep per day is a useful measure. A decrease in efficiency in performing psychological tests early in addiction, followed by a return to the control level later in addiction, also serves as a measure of tolerance. If the drug produces changes in the electroencephalogram, a return to the normal pattern is very good evidence of tolerance. The development of tolerance to the pain threshold-elevating action of the drug is followed by repeating the Hardy-Wolff measurements at intervals throughout addiction. Tolerance to the emetic effects is easily observed clinically, and tolerance to the depression of appetite is manifest by a return of the caloric intake to the normal level.

Detection of Emotional Dependence. Good evidence of habituation may be obtained by careful observation of the behavior of the patients. If they express satisfaction with the effects of the drug, stop nearly all productive activity, neglect their persons and their quarters, and spend a great deal of time in bed in a semi-somnolent state, the drug under test has considerable habituation liability. If changes in psychological tests similar to those which occur during morphine addiction are found, the clinical evidence is greatly reinforced. If the subjects like the effects of the drug they will request increases in the dosage level throughout addiction. They will even beg for more of the drug when definite evidence of toxic effects are present. If the subjects fail to get satisfaction from the drug, they will usually ask for increases in dosage early in addiction in the hope that more of the drug will produce the effects they desire. If, after several increases in dosage, they have not experienced sufficiently pleasurable effects, they will not ask for further increases but simply tolerate further elevations as part of a bad bargain. Under such circumstances, some of the subjects will ask for, or even demand, termination of the experiment. Important evidence of habituation liability of the drug may be obtained after the addiction experiment is completed. If the men continue to ask for the drug for weeks and months after it has been withdrawn, and if they persist in hatching schemes to obtain it, the drug undoubtedly has habituation liability. The value that addicts place on the drug, relative to the value they place on morphine, can be assessed roughly by arranging experiments in which they are offered a choice of either morphine or the new drug. If a number of patients repeatedly choose 10 to 30 mg. of the new drug in preference to 30 mg. of morphine, the drug has high habituation liability. If they do not choose the new drug until 120 to 240 mg. are offered in exchange for 30 mg. of morphine, the drug has low habituation liability.

Detection of Physical Dependence. Physical dependence on the new drug is studied by completely and abruptly withdrawing the drug. Once the drug has been withdrawn, no narcotic drugs whatever are given until the observations are complete. During the withdrawal period, the study subjects are completely isolated from the remainder of the population of the institution, and special precautions are taken to prevent the introduction of contraband narcotics into the experimental ward. Observations for signs of abstinence are made as described under "Substitution for Morphine." If clear-cut evidence of physical dependence is not obtained after a short period of addiction—2 months or less—the experiments are repeated using an addiction period of 6 months or more.

Results of Testing Addiction Liability of Drugs in the Methadone Series.

In sufficient dosage, racemic methadone was shown to produce intense euphoria in former morphine addicts which was manifested by increased talkativeness, boasting, requests for more of the drug, and, with large doses, marked sedation.⁶ The effects of methadone on psychological tests were similar to those of morphine. The euphoria produced by methadone came on more slowly than the euphoria produced by morphine, but persisted much longer. Intravenous injections of methadone produced striking euphoria, which experienced addicts described as similar to that following intravenous injections of heroin or Dilaudid. After intravenous injections of methadone, the addicts would writhe in joy, and say, "Oh, boy! that's a good shot. What is the name of that dope, can you get it outside? Will it be put under the Law? If God made anything better than that, He kept it for Himself." In blind experiments, experienced addicts could not distinguish the subjective effects of methadone from those of morphine, heroin, or Dilaudid. Some men who were experimentally addicted to methadone came to prefer it to morphine, and in subsequent experiments would exchange morphine for methadone on a weight-for-weight basis. The optical isomers of methadone were quite different in their effects. Dextromethadone did not produce detectable euphoria even when 120-mg. doses were given. Fifteen mg. of levomethadone⁷ produced intense euphoria. Racemic isomethadone⁷ (6-dimethylamino-5-methyl-4,4-diphenyl-3-hexanone) produced definite euphoria in doses of 30 to 60 mg. Methadol (6-dimethylamino-4,4-diphenyl-3-heptanol), the alcoholic analogue of methadone, did not produce euphoria.

During experimental addiction to racemic methadone, tolerance developed to the sedative, emetic, miotic, and pain threshold-elevating actions^{8,9} (FIGURE 1). Similar evidence of tolerance was obtained during experimental addiction to racemic isomethadone.⁷

The behavior of men addicted to methadone was similar to the behavior seen during morphine addiction.^{8,9} The patients ceased all productive activity, neglected their persons and their quarters, and spent most of their time in bed in a semi-somnolent state which they regarded as very pleasurable. Psychological changes during addiction to methadone were

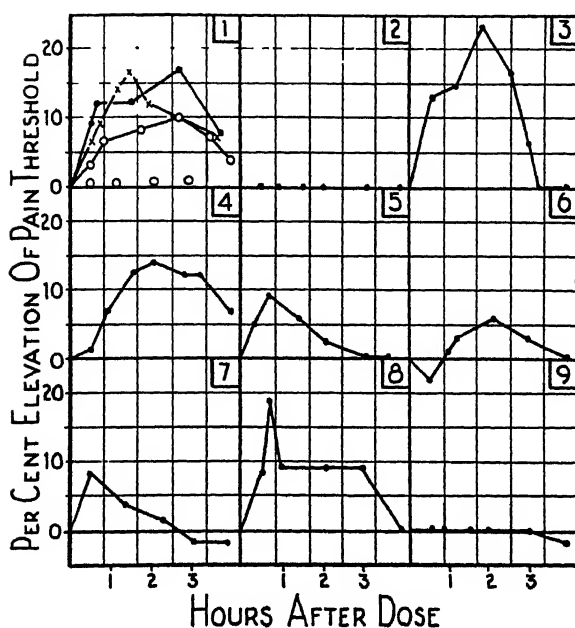


FIGURE 1. Development of tolerance to the pain threshold-elevating action of methadone in Subject 729. *Curve 1:* Solid circles, greatest response to a 5-mg. dose of methadone prior to addiction. Open circles, smallest response to a 5-mg. dose. Crosses, average response to a 5-mg. dose. Open circles with dots, control after injection of distilled water. *Curve 2:* Response to a 5-mg. dose after 7 days of addiction. *Curve 3:* Response to a 15-mg. dose on the 8th day of addiction. *Curve 4:* Response to a 15-mg. dose after 14 days of addiction. *Curve 5:* Response to a 15-mg. dose after 21 days of addiction. *Curve 6:* Response to a 25-mg. dose after 28 days of addiction. *Curve 7:* Response to a 30-mg. dose after 35 days of addiction. *Curve 8:* Response to a 45-mg. dose after 42 days of addiction. *Curve 9:* Response to a 45-mg. dose after 56 days of addiction.

similar to those seen during morphine addiction.^{8,9} During addiction to methadone, the patients continually requested increases in dosage. The behavior of men experimentally addicted to isomethadone was somewhat different. In the first week of addiction, the subjects were exhilarated and slept poorly.⁷ As the dosage was elevated, this effect disappeared and the men showed evidences of mild sedation, which was never so marked as with morphine or methadone. Some of the subjects continued to work throughout addiction to isomethadone. At first, the patients requested increases in dosage and complained that the effects were not as pleasant as they had hoped. Later in the experiment, they did not ask for any further increases in dosage and simply seemed to tolerate the elevations in the amount of the drug as a necessary evil. Following recovery after withdrawal, only 2 of the 10 subjects took isomethadone in preference to morphine, and then only when 240 mg. of isomethadone were offered in exchange for 30 mg. of morphine.

Physical Dependence on Drugs of the Methadone Series. Racemic meth-

adone⁷⁻⁹ and levomethadone⁷ produced a spectacular decrease in the intensity of abstinence when administered to men, 28 to 32 hours after withdrawal of morphine (FIGURE 2). Racemic methadone was approxi-

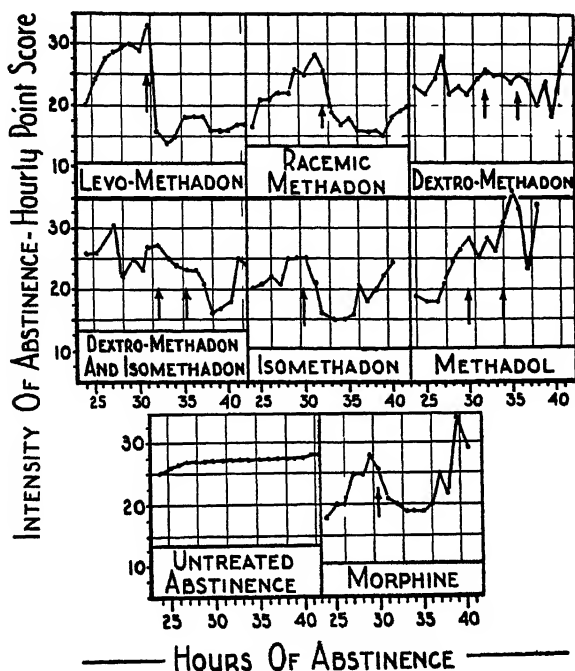


FIGURE 2. Effects of drugs of methadone series on abstinence from morphine. Ordinates: Average intensity of abstinence expressed in hourly points.⁴ Abscissae: Hours of abstinence. Arrows indicate injection of drugs. 10 subjects received levomethadone, racemic methadone, and isomethadone; 4 received 2 doses of dextromethadone; 6 received one dose of dextromethadone followed by one dose of isomethadone; 4 received one dose of morphine. Course of untreated abstinence taken from data of HIMMELSBACH.⁴

mately 4 times as potent in relieving abstinence as was morphine, and levomethadone was 8 times as effective as morphine. The effects of both racemic methadone and levomethadone on the intensity of abstinence persisted much longer than did the effects of morphine. Racemic isomethadone (6-dimethylamino-5-methyl-4,4-diphenyl-3-hexanone) lowered the intensity of abstinence from morphine but was not as effective as morphine in this respect. Dextromethadone⁷ and methadol⁷ (6-dimethylamino-4,4-diphenyl-heptanol) had no effect on the morphine abstinence syndrome.

Racemic methadone^{8,9} (FIGURE 3) and levomethadone⁷ (FIGURE 4) completely prevented the appearance of signs of abstinence when substituted for morphine in cases addicted to large doses of that drug. Following abrupt withdrawal of racemic methadone^{8,9} or levomethadone⁷ after substitution for morphine, an abstinence syndrome developed which was quite characteristic in its course and manifestations. No signs were noted

and the patients made no complaints until the 3rd or 4th day of abstinence. Thereafter they complained of fatigue, weakness, anxiety, vague

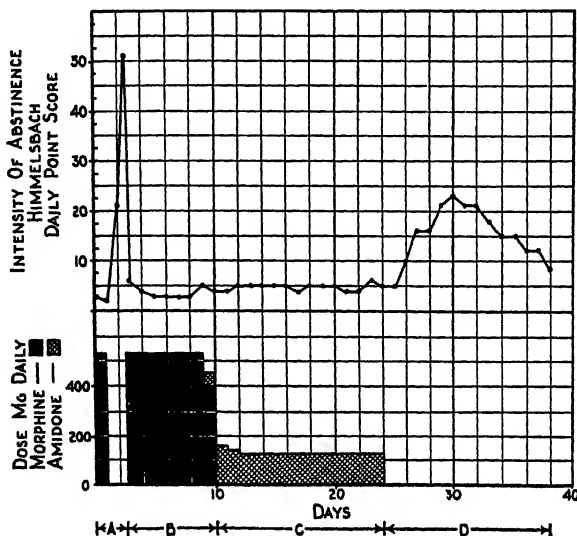


FIGURE 3. Substitution of methadone for morphine, average of 12 cases. A, preliminary withdrawal from morphine. B, restabilization on morphine. C, substitution of methadone. D, withdrawal of methadone.

abdominal discomfort, and "hot and cold flashes." Objectively, signs of disturbances in autonomic function were much less prominent than after the withdrawal of morphine. Body temperature, pulse and respiratory rates, and systolic blood pressure were increased. Caloric intake was consistently depressed. The average intensity of abstinence following withdrawal after substitution for morphine did not exceed 21 daily points (during a preliminary test period of abstinence from morphine, the same patients scored an average of 52 daily points from the 24th to the 36th hour of withdrawal). The onset of abstinence from methadone was slower than the onset of abstinence from morphine, and the course was somewhat prolonged as compared to the course of abstinence from morphine.

Isomethadone⁷ partially suppressed signs of abstinence when substituted for morphine at a ratio of 1 mg. of isomethadone for each 1.33 mg. of the substitution dose of morphine. Minor signs of abstinence appeared on the 2nd day of substitution and persisted through the 5th day of substitution (FIGURE 4). Following abrupt withdrawal of isomethadone after 10 days' substitution for morphine, an abstinence syndrome was detected 12 hours after the last dose of isomethadone was administered. Qualitatively, abstinence from isomethadone was similar to abstinence from morphine. Many signs of disturbed autonomic function—yawning, lacrimation, rhinorrhea, mydriasis, etc.—were noted. The patients were restless,

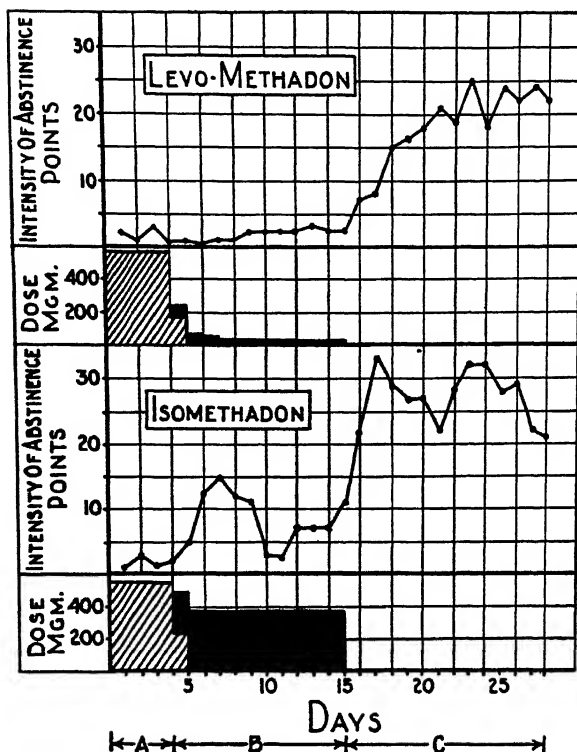


FIGURE 4. Substitution of levomethadone and isomethadone for morphine. Intensity of abstinence is expressed in daily points.² A, period of stabilization on morphine. B, period of substitution of isomethadone for morphine. C, period of abstinence.

had poor appetites, vomited, and showed slight elevations in rectal temperature, pulse rate, respiratory rate, and systolic blood pressure. Quantitatively, the intensity of abstinence was about equal to abstinence from codeine or meperidine. The Himmelsbach daily point score rose to 33 points on the 2nd day of abstinence. Abstinence from isomethadone was more intense than abstinence from methadone, but less intense than abstinence from morphine.

Following abrupt withdrawal of racemic methadone^{8, 9} from patients who had had their dosage elevated to 180 to 400 mg. of drug daily over the course of 2 to 6 months' experimental addiction, an abstinence syndrome developed which was qualitatively identical with that which ensued after withdrawal of methadone following substitution for morphine (*vide supra*). The intensity of the abstinence was more severe in the patients who were directly addicted to the drug for 6 months than in patients who had the drug substituted for morphine.^{8, 9} The average intensity of abstinence rose to 33 points on the 9th day of abstinence and was still above 20 points on the 14th day after withdrawal (FIGURE 5).

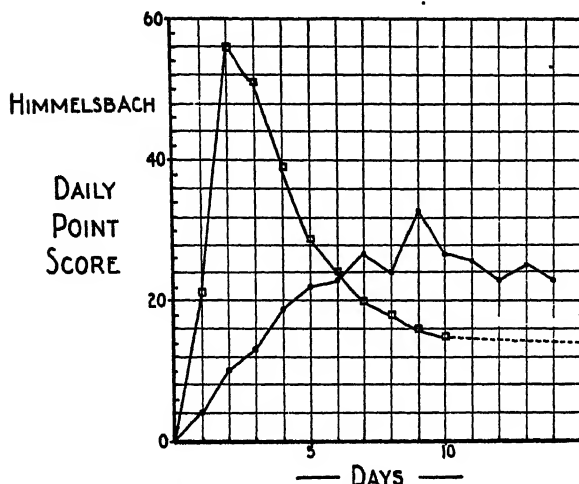


FIGURE 5. Intensity of abstinence after administration of methadone for four and one-half to six months. Line broken by solid circles: days after withdrawal of methadone, average of 5 cases. Line broken by squares: days after withdrawal of morphine, average of 65 cases.

After abrupt withdrawal of isomethadone from 10 patients who received doses of isomethadone increasing to 240 to 360 mg. daily during 6 to 8 weeks' experimental addiction, an abstinence syndrome appeared within 12 to 18 hours, which was similar to the abstinence syndrome following withdrawal of isomethadone after substitution for morphine (FIGURE 6). Qualitatively, the isomethadone abstinence syndrome was quite similar to the morphine abstinence syndrome. Quantitatively, the intensity of abstinence from isomethadone was about equal to that of abstinence from codeine.

The results show that racemic methadone and levomethadone have all the characteristics of addicting drugs. In sufficient dose, they produce intense euphoria in former morphine addicts. The habituation liability of these drugs is at least equal to, and perhaps greater than, that of morphine. Definite evidence of physical dependence was observed following withdrawal of the drugs after substitution for morphine, or after 2 to 6 months' addiction to racemic methadone. Although the physical dependence liability of methadone is less than that of morphine, the emotional dependence liability of the drug is so great that methadone should be regarded as being almost as dangerous as is morphine.

Isomethadone is also an addicting drug. Tolerance develops to many of its actions. It produces definite euphoria in former morphine addicts. Both the habituation and physical dependence liability of isomethadone are less than those of morphine. The total addiction liability of isomethadone seems to be roughly equivalent to that of codeine.

Discussion. The choice of methods to be used in testing a particular

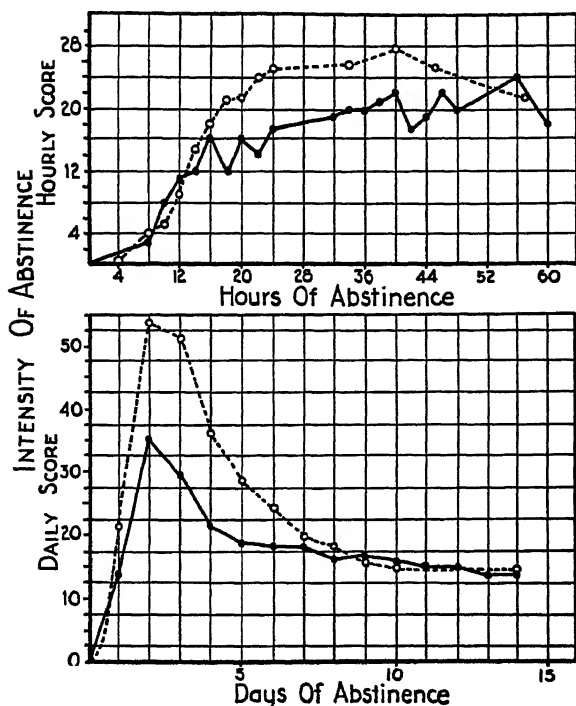


FIGURE 6. Intensity of abstinence following withdrawal of isomethadone after experimental addiction for 42 to 59 days, averages of 10 cases. *Upper curve:* onset of abstinence from isomethadone as compared to onset from morphine. Intensity expressed in hourly points.¹ *Lower curve:* course of abstinence from isomethadone compared to abstinence from morphine. Intensity of abstinence expressed in daily points.² Solid line shows intensity of abstinence from isomethadone. Dotted line shows intensity of abstinence from morphine based on 65 cases of KOLB & HIMMELSBACH.²

drug depends on the amount of the drug available, and on whether or not it is a member of a series with known addiction liability. If only a small amount of the drug is available, the single dose and/or the substitution methods must be used. If the drug is a member of a known addicting series, the single dose methods usually yield sufficient information. If the drug relieves physical dependence and produces euphoria, it has addiction liability. If the drug does not relieve physical dependence, substitution studies and direct addiction experiments must be carried out.

When the drug to be tested is a member of a completely new series of analgesic drugs, all the methods must be used. A drug should never be regarded as non-addicting unless patients have received large doses of the compounds for at least 6 months.

The final test of the addiction liability of a new drug comes when the drug is released for sale and is widely used in clinical practice. Meperidine,⁵ which Himmelsbach judged to have addiction liability on the

basis of the results of tests identical with those described above, has been available for 5 years. During that period, it has been widely abused by morphine addicts, and, since 1946, about 20 meperidine addicts who were never addicted to morphine have been admitted to the U. S. Public Health Service Hospital at Lexington, Kentucky. Most of the primary meperidine addicts which have been studied at Lexington were originally given meperidine for the relief of pain or asthma. This suggests that, even though our experiments are designed to simulate conditions of abuse, they are not without value in predicting addiction liability under conditions of legitimate medical use. Methadone has been available for only a year, but already morphine addicts are abusing the drug. Several patients who were using both morphine and methadone have entered the Lexington hospital in recent months. These cases support the validity of the methods of testing addiction liability described above.

Summary

1. In assessing the addiction liability of a new analgesic drug, attention should be given to all three of the characteristics of addiction to the opiate drugs—tolerance, physical dependence, and habituation (emotional or psychic dependence). Physical dependence and emotional dependence are equally important. The final tests of addiction liability must be carried out on human subjects. Addiction-liability tests which simulate conditions of abuse are more important than tests of addiction liability under conditions of medical use.
2. Four methods of testing addiction liability are described: administration of single doses for the detection of euphoria; the effect of single doses on abstinence from morphine; substitution of the new drug for morphine in cases strongly addicted to morphine; and direct addiction to the new drug.
3. Racemic methadone, levomethadone, and racemic isomethadone are addicting drugs. The total addiction liability of racemic methadone (or levomethadone) is nearly equal to the addiction liability of morphine. The addiction liability of racemic isomethadone is about equal to that of codeine.

References

1. HIMMELSBACH, C. K., & L. F. SMALL. 1937. Clinical studies of drug addiction. II. "Rossium" treatment of drug addiction with a report on the chemistry of "rossium." Pub. Health Rep., Suppl. 125: 1-18.
2. KOLB, L., & C. K. HIMMELSBACH. 1938. Clinical studies of drug addiction. III. A critical review of the withdrawal treatments with method for evaluating abstinence symptoms. Am. J. Psychiat. 94: 759.
3. HIMMELSBACH, C. K. 1939. Studies of certain addiction characteristics of (a) dihydromorphine ("paralimorph"), (b) dihydrodesoxymorphine-D ("desomorphine"), (c) dihydrodesoxycodine-D ("desocodine"), and methyl dihydromorphinone ("metopon"). J. Pharm. & Exp. Therap. 67: 239-249.

4. HIMMELSBACH, C. K., & H. L. ANDREWS. 1943. Studies on modification of the morphine abstinence syndrome by drugs. *J. Pharm. & Exp. Therap.* 77: 17-23.
5. HIMMELSBACH, C. K. 1942. Studies on the addiction liability of "demerol" (D-140). *J. Pharm. & Exp. Therap.* 75: 64-68.
6. ISBELL, H., A. J. EISENMAN, A. WIKLER, & K. FRANK. 1948. The effects of single doses of 6-dimethylamino-4-4-diphenyl-3-heptanone (amidone, methadon, or '10820') on human subjects. *J. Pharm. & Exp. Therap.* 92: 83-89.
7. ISBELL, H., & A. J. EISENMAN. 1948. The addiction liability of some drugs of the methadon series. (In press.)
8. ISBELL, H., A. WIKLER, N. B. EDDY, J. L. WILSON, & C. F. MORAN. 1947. Tolerance and addiction liability of 6-dimethylamino-4-4-diphenyl-heptanone-3 (methadon). *J. A. M. A.* 135: 888-894.
9. ISBELL, H., A. WIKLER, A. J. EISENMAN, M. DAINGERFIELD, & K. FRANK. 1948. The addiction liability of 6-dimethylamino-4-4-diphenyl-3-heptanone (methadon, amidone, or '10820') in man. (In press.)

THE IMPORTANCE OF ADDICTION TO THE NEWER SYNTHETIC ANALGESICS IN HUMAN THERAPY

By ROBERT C. BATTERMAN

*Department of Therapeutics,
New York University College of Medicine, New York, N. Y.*

WITH the introduction of new synthetic analgesic agents, considerable attention has been focused on the possibility of the occurrence of addiction with their use. Although it appears that some of these drugs possess addictive properties, insufficient data have accumulated, when used in bona fide medicine, attesting to the relative addiction liability of these compounds as compared to known opiates. Any consideration of the addiction liability of the new analgesics must, therefore, include a discussion of the problem of addiction in general.

If by the term "addiction" is meant the abuse of a drug by reason of excessive use and continued administration, then it is a well-known clinical fact that every medication, regardless of its therapeutic merit or potency, may in susceptible individuals be classified as an addictive drug. Obviously, there are many objections to this definition. Emphasis is placed upon "susceptible individuals" rather than upon any particular drug. Abuse of any medication, to satisfy a psychic need, should be primarily labeled as habituation. In contrast to this, if a drug produces upon chronic administration an alteration of the physiological functions of the body so that its withdrawal results in typical objective physical abstinence phenomena, then such a state is classified as addiction. The definitions distinguishing habituation and addiction are also inadequate, for once again emphasis is placed on only one facet of the problem, in this case the pharmacological effects of the drug. It is apparent that drug addiction is a problem of the individual as well as of the drug. It may be advantageous, in view of the confusion and controversies resulting from adherence to any definition of addiction, to obtain a new term which would more adequately express the dual concept of the problem. It may be possible, in this manner, to avoid the widespread misuse by both the medical profession and the lay public of the term "addiction," or the labeling of an individual as an "addict" in the absence of clear-cut evidence. It is of interest to note that the words "addiction" or "addict" were rarely employed prior to 1910 and that these terms were not used in the narcotic laws enacted by Congress until 1929, at which time the narcotic farms for the confinement and treatment of persons addicted to "habit-forming narcotic drugs" were established. There is absolutely no question that legal restriction of these drugs is essential, but it is unfortunate that a new connotation that "an addict is a criminal"

has been implanted upon habitual users of these drugs. Although many criminals, because of personality disorders, resort to narcotic drugs, it definitely beclouds the issue to consider all individuals who abuse drugs as criminals.

In the bona fide use of these drugs in medicine, the physician is often in fear of the legal aspects involved in prescribing narcotics. As Kolb and DuMez¹ point out, "the numerous reports and forms which physicians are required to make out in order to prescribe narcotics in any form tends to keep them alert to the dangers of these drugs." There is no doubt that the legal safeguards set up have played a considerable role in decreasing the number of medically addicted patients. However, from the viewpoint of the patient who actually requires these medications, it makes one wonder if treatment may have been inadequate because the physician's concern is first with legal forms and second with the needs of the patient. I do not wish to imply that the physician should overlook the dangers of addiction to narcotic drugs, but in the use of these drugs in medicine the likelihood of addiction resembling the type exemplified by the psychopathic individual is small. It is true that promiscuous administration of these drugs by some physicians has resulted in patients continuing their use as addicts. However, this group in comparison to the overall census of addicts is also small. Suffice it to state now that most physicians are, unfortunately, confused in regard to the addictive liability of these drugs and do not distinguish between the underlying psychogenic component of the patient and the pharmacological effects of the drug.

Habituation is possible with every known drug, including milk sugar and sodium chloride as well as those drugs exerting a potent action upon the higher brain centers. In my experience, it is not an unusual occurrence for patients to require a placebo repeatedly during the day, or whenever the need arises to sustain or achieve a state of well-being. Cessation of such therapy, particularly if the original reason for the medication continues, results in excitement, restlessness, nervousness, and incessant demands for the renewal of the medication. With the exception of typical abstinence phenomena, these patients, from the subjective point of view and psychic reactions, are indistinguishable from individuals with true addiction. Unfortunately, a patient with psychic addiction very often becomes a problem in management. The situation is particularly disturbing if the patient, because of a disease such as bronchial asthma or angina pectoris which possesses marked psychosomatic components, feels that any moment will be his last. Many of these patients are classified as addicts mainly because they are very insistent about obtaining medication. Unless physical dependence to the drug can be demonstrated, the physician is guilty of a serious injustice to the patient if he makes such a classification.

It is generally agreed that the fundamental aspect of addiction revolves about the underlying psychic deficiency of the individual. Thus,

Fenichel² states that "the origin and nature of the addiction are not determined by the chemical effect of the drug but by the psychological structure of the patient." Kolb's³ classification of drug addicts is verification of this fact.

An individual possessing a pre-morbid personality will grasp unconsciously at any means to satisfy the psychic void and thus obtain a sense of security or maintenance of self-esteem. The use of a potent drug is only incidental in the overall picture of addiction. It enters into the picture only when opportunity presents itself for use. In light of this, it is immaterial whether an individual resorts to alcohol, cocaine, morphine, or any other drug so long as this psychic void is filled. It is of interest to note that some patients resort to excessive intake of food to satisfy their psychic needs while, in other cases, the mere insertion of a needle in the so-called "needle addicts" is sufficient. However, the problem goes further than this because certain drugs have the ability to alter the physiological functions of the body, which results in physical dependence. In the pathogenesis of addiction, an individual usually goes through a phase of habituation. As the drug is continued, the individual begins to experience increasingly severe symptoms when the drug is not taken. It is almost immediately learned that repetition of the drug prevents or abolishes these unpleasant symptoms. A vicious cycle is thus established with the individual unable to discontinue the drug at will. It is this phase of drug addiction which makes the drug addict a problem to himself as well as to the community.

The addiction liability of the new analgesic agents must, therefore, be considered in two phases: first, the liability of the individual to use these new preparations and, second, the liability of the drug to produce physical dependence.

In 1943, Himmelsbach and I⁴ indicated that, in terms of likelihood of developing an addiction to a new analgesic drug, there exist four types of individuals or conditions which must be considered independently. At that time, we worked with Demerol, knowing that this drug can produce physical dependence. The same four categories can be utilized for all new analgesic drugs.

First, we have the use of these drugs by former opiate addicts. It has been definitely established that Demerol, methadone, and several derivatives of the latter can result in physical dependence when administered to former morphine addicts. Unrestricted use of Demerol and methadone, the two currently available synthetic drugs, in this group of individuals leads to abuse and invariably to an addiction indistinguishable, for the most part, from that noted with the opiates used previously. Correctly interpreted, this type of addiction should be classified as secondary or substitution addiction. It is important to take this into consideration because addicts develop physical dependence to these drugs with an ease not noted in patients receiving the same drugs in the bona fide use in medicine. There is a possibility that previous opiate use has altered the

physiological background so that the individual becomes more susceptible to the physical dependence effects of these drugs. An analogy of this phenomenon may be found in the studies with desomorphine upon monkeys by Eddy and Himmelsbach.⁵ "The abrupt withdrawal of dihydrodesoxymorphine-D, the administration of which was superimposed upon a previous morphine addiction, resulted in the appearance of abstinence manifestations not significantly distinguishable from those appearing after abrupt withdrawal of morphine alone, and in marked contrast to the very slight modifications of behavior which occurred upon abrupt withdrawal of dihydrodesoxymorphine-D when it was the only drug used."

Whether an opiate addict will resort to these newer analgesic agents for control of his withdrawal symptoms depends upon several factors. These include the degree of "lift" obtained by the drug, or the ability of the medication to restore emotional normalcy; the ability of the drug to control the signs and symptoms of abstinence in terms of dose, tolerance, and duration of effect; and the availability of the medication through illicit channels. With rare exception, the confirmed morphine addict is not satisfied by the mild euphoric action of Demerol. Add to this the large dose usually required for control of withdrawal symptoms, the rapid development of tolerance, the progressively decreasing interval between "shots," and the difficulty of obtaining Demerol as compared to opiates, it then becomes evident why Demerol is not the drug of choice by this group of individuals. Some addicts will occasionally use Demerol when unable to obtain morphine. However, the use of Demerol to the exclusion of morphine, or one of the other opiate drugs, by an individual previously addicted to them is very unusual. According to the data graciously supplied to me by Drs. Vogel and Chapman⁶ of the United States Public Health Service Hospital at Lexington, Kentucky, such individuals constitute about one-half of one per cent of the total number of addicts admitted to that institution during the past year.

It is too soon to evaluate methadone in this regard. However, it is well known that methadone and several of its derivatives are satisfactory agents for controlling the withdrawal symptoms of morphine when given to previous morphine addicts. Several patients, other than those experimentally studied, have already been treated for methadone addiction at the Lexington, Kentucky, hospital, and I recently had the opportunity of observing a similar patient who substituted Demerol and then methadone for his addiction to morphine. In view of the fact that methadone produces in such individuals greater satisfaction than does Demerol, it is likely that a greater number of secondary addicts will appear in the near future.

The possibility of primary addiction in an unstable person who otherwise would abuse an opiate to the point of addiction, but who has had no previous opiate experience, constitutes the second and third categories. In one group, we have individuals who make original contact with the drug as a result of its administration by a physician for a defi-

nite illness, and because some psychic deficiency is alleviated, the drug is continued and abused. In the other group, we have physically healthy individuals who, instead of using an opiate for pleasure and curiosity, resort to the newer analgesic drug. There is no doubt that primary addiction has been noted in the first instance, but, so far as I know, no case has been reported as being primarily addicted to Demerol or methadone because these drugs were used first in the same way morphine or heroin is used by the typical psychopathic individual. Several reasons for this come to mind and include the relatively small degree of euphoria obtained as contrasted with morphine and heroin; the high probability of acute toxicity and unpleasant untoward reactions frequently noted when these new drugs are administered parenterally, particularly if the individual is ambulatory; and, finally, the difficulty of obtaining the medication through illicit channels.

Let us return for a moment to the group of medically addicted patients who possess all the psychic attributes conducive to addiction. According to the Public Health Service data from Lexington, Kentucky, this group using Demerol constitutes one-half of one per cent of all addicts. This is approximately one-sixth the incidence of similar patients addicted to morphine.

Reports in the literature regarding this phase of addiction to Demerol are very scanty. The first reference⁷ to primary addiction with Demerol resulting from medical use appeared September, 1946, a period of approximately 7 years following the introduction of the drug and 3 years following availability of the drug in this country. Commissioner Anslinger of the Federal Narcotic Bureau published abstracts of reports of 15 individuals addicted to Demerol. Of this group, 7 cases were of the secondary type and do not enter into this discussion. In case No. 5, no data other than a statement that "he was reported to be addicted to demerol" was furnished. In the ninth case, the patient was labeled an addict merely because she demanded the drug to relieve her pain. Patient No. 14 was stated to have a "craving for Demerol" without indication that this may have been habituation rather than physical dependence. In the eighth patient, there was a history of opiate use, but not addiction prior to Demerol abuse. The remaining four cases were physicians, two of whom, admittedly, had also had some opiate previously, while one had a chronic disease which would make it very likely that some opiate had been used. Thus, in only one case could the diagnosis of primary addiction be accepted without doubt. Wieder⁸ from the United States Public Health Service Hospital of Lexington, Kentucky, in December, 1946, reported three cases of Demerol addiction. One case was definitely a secondary type. Another case was a nurse who had access to opiates and used these preparations for over 20 years before resorting to Demerol. The third patient also had considerable experience with codeine and morphine prior to the use of Demerol, although at no time was he considered to be an opiate addict.

These two reports are the only available evidence in the literature regarding primary addiction to Demerol. Correspondence with the Lexington, Kentucky, hospital also indicates that there have been additional patients of this type. It is my impression that the majority of these patients (the exact number is unavailable to me) have had previous opiate administration but have not been classified as opiate addicts because they gave no history of abstinence manifestations with their use. Classification of these patients as true primary Demerol addicts is therefore open to question. We have insufficient knowledge of the action of opiates to exclude the possibility that previous administration of these preparations does not alter physiological or chemical functions short of demonstrable physical dependence. Again, the experiments of Eddy and Himmelsbach with "desomorphine" strengthen this viewpoint. Primary addiction to Demerol undoubtedly does occur, but the evidence available to date requires further careful evaluation.

The fourth category is the development of physical dependence "in an otherwise normal person suffering an illness requiring prolonged administration of a potent analgesic." The addiction liability of morphine in this group of patients was well presented by Lee.⁹ In all cases, typical abstinence phenomena occurred following parenteral and oral use for a period of chronic administration of 21 days. This is in accord with clinical experience that patients receiving morphine for a chronic illness will, with rare exception, develop physical dependence to the drug within two to four weeks of therapy. The psychic component, as a rule, is not prominent in this group of patients and, because of this, the patient is not considered an addict. However, in view of the physical dependence to morphine, these patients, according to the definition of addiction, should be so classified. The addiction liability of the new analgesic drugs in terms of duration of administration to achieve physical dependence in a similar group of patients is very low. In reference to Demerol, I was able to follow 47 patients on continuous administration both orally and parenterally for many weeks or months, and in a few instances for over a year, without noting any signs or symptoms of abstinence upon withholding the drug. Curry¹⁰ has reported a similar case. The Public Health Service Hospital has also not encountered any accidental primary addiction to Demerol in a normal individual.

In regard to methadone, information about the last three categories are lacking because the drug has not had sufficient clinical use. It may be predicted, however, that primary addiction will be noted with extended use, but, in all probability, because of its low addiction liability, the incidence of its occurrence will be small.


The final point for discussion is the danger involved, other than addiction, from the chronic administration and abuse of these drugs. Excessive doses of Demerol result in cerebral irritability that manifests itself by uncoordinated muscle movements, tremors, and in the extreme state by convulsions. Toxic psychosis may occur. The atropine effects of

Demerol result in visual disturbances and marked dryness of the mucous membranes. Prolonged use of methadone may also result in serious cumulative toxicity. In a group of 33 patients reported by Oshlag and me,¹¹ mental confusion and toxic psychosis occurred in 37.5 per cent of patients over the age of 50 who received the drug in therapeutic doses of 2.5 to 5.0 mg. repeatedly for 5 days or longer. Data on methadone addicts are not available, but it is anticipated that this toxic manifestation will be serious in view of the usual practice of these individuals to exceed the therapeutic dose. The one methadone addict I had the opportunity to observe demonstrated mental deterioration and memory lack which subsided when the drug was discontinued.

In summary, the addiction liability of the new analgesic drugs must be considered from the point of view of the individual as well as the drug. The major problem in addiction is the psychic structure of the individual. The new analgesic drugs produce physical dependence. To date, this has been satisfactorily demonstrated in previous opiate addicts or individuals who have had previous opiate experience. Primary addiction undoubtedly occurs, but the available evidence is unsatisfactory. The overall incidence of addiction to Demerol is small considering the widespread use of the drug in medicine. Conclusions regarding methadone must be held in abeyance until it has had further clinical trial.

Bibliography

1. KOLB, L., & A. G. DuMEZ. 1924. The Prevalence and Trend of Drug Addiction in the United States and Factors Influencing it. U. S. Pub. Health Rept. 924.
2. FENICHEL, O. 1945. The Psychoanalytic Theory of Neurosis. W. W. Norton & Co. New York.
3. KOLB, L. 1925. Types and characteristics of drug addicts. Mental Hygiene 9: 300-313.
4. BATTERMAN, R. C., & C. K. HIMMELSBACH. 1943. Demerol—a new synthetic analgesic. J. A. M. A. 122: 222.
5. EDDY, N. B., & C. K. HIMMELSBACH. 1936. Experiments on the Tolerance and Addiction Potentialities of Dihydrodesoxymorphine-D ("Desomorphine"). U. S. Pub. Health Rept., Suppl. 118.
6. VOGEL, V. H., & K. W. CHAPMAN. Personal communication.
7. ANSLINGER, H. J. 1946. Correspondence to the Editor. J. A. M. A. 132: 93.
8. WIEDER, H. 1946. Addiction to meperidine hydrochloride (Demerol hydrochloride). J. A. M. A. 132: 1066.
9. LEE, L. E., JR. 1942. Studies of morphine, codeine and their derivatives XVI. Clinical studies of morphine, methyl dihydromorphinone (Metopon) and dihydrodesoxymorphine-D. J. Pharm. & Exp. Therap. 75: 161.
10. CURRY, J. J. 1947. Habituation to meperidine hydrochloride (Demerol hydrochloride). J. A. M. A. 133: 243.
11. BATTERMAN, R. C., & A. M. OSHLAG. 1948. The effectiveness and toxicity of methadon. Fed. Proc. 7: 1.

56377


tive application if in the light of further knowledge it develops that the drug is not safe for use. Details of these procedures need not be discussed here.

Comment may be offered on two points mentioned above. In the first place, making a new drug application effective does not constitute either endorsement or approval of a new drug. Efficacy for the purpose intended is not a primary consideration in the review of the application. This review has to do with the evidence for safety. However, efficacy may come into the picture. A hypothetical case may illustrate this. If thiouracil had, perchance, been submitted simultaneously or subsequently to the submission of propylthiouracil, the two substances being essentially equivalent so far as therapeutic efficacy is concerned, the greater toxicity of thiouracil would undoubtedly have resulted in a refusal to permit the application for thiouracil to become effective.

A second comment has to do with the brochure in which, in the case of the majority of new drugs, the directions for use are stated. It is the practice of the New Drug Section to give very careful consideration to this material. It is here that the introducer can set down the information which the physician requires for the effective and *safe* use of the material. It is in the brochure that attention can be called to the side effects, or the effects of overdosage, and to the circumstances under which a maximum of safety can be attained. Actually, the very fact of newness means that standard texts on drugs and drug action and uses which are readily available to the physician are not likely to contain the required information.

No word has been said, up to this point, regarding new analgesic drugs. All of the requirements with respect to any new drug of course apply to a new analgesic agent. A new drug application covering a drug which is introduced for use in the relief of pain must be judged by the same fundamental standards, evaluated with respect to the same points, as other potent agents. How does it act? What are the frequency and severity of the side reactions? What is the margin of safety? What are the possibilities or likelihood of misuse?

The answer to the last question is, of course, the one of greatest importance. As the reader is aware, there are a number of substances which have therapeutic value and which should not be denied the physician, but which nevertheless can readily be misused. Amphetamine, the barbiturates, bromides, alcohol, cocaine, opium and its derivatives, all are subject to abuse, and these abuses may become major social problems. It is the position of the Food and Drug Administration, as already stated, that maximum safety is attained when the physician is fully informed as to all of the hazards of use. It is, therefore, the belief of the Administration, and applicants are so advised, that applications for drugs having analgesic properties should include reports of sufficient work to establish clearly just what hazards of habit formation or of addiction may exist. More than this, such information should be adequate

to establish the relative degree of hazard as compared with that from other drugs of equal effectiveness. It is not the responsibility of the Food and Drug Administration to say who should make such studies. It *is* its responsibility to be certain that these are adequate, both qualitatively and, in so far as is possible, quantitatively. Only in the light of such information is a maximum of safety possible, and only with such information before it can the Food and Drug Administration reach a sound and proper conclusion with respect to the safety for use of a new analgesic.

REGULATORY PROBLEMS OF THE NEW ANALGESICS UNDER THE NARCOTIC LAW

By H. J. ANSLINGER

Bureau of Narcotics, Treasury Department, Washington, D. C.

FROM the standpoint of the application of drug traffic control measures under Federal narcotic laws in effect prior to July 1, 1944, the new analgesics could be divided into two classes; (1) new analgesic opium derivatives and (2) synthetic analgesics. In the case of new analgesic opium derivatives, of which Metopon is an example, no new Federal legislation was required to make them subject to the existing Federal laws relating to narcotic drugs. Thus, Metopon, being a morphine derivative, automatically became subject to the same control as morphine sulphate, in so far as the application of the Federal narcotic drug laws is concerned. An additional measure of control over the production, distribution, and use of Metopon was afforded through the instrumentality of a patent for the drug which was assigned to, and is now owned by, the government.

I believe it is now generally understood that Metopon was found to possess certain advantages over other opium derivatives, currently in medical use, in the treatment of the chronic suffering of malignancies. Therefore, the distribution of Metopon was limited by patent license agreement to a single medical purpose: the oral administration for chronic pain relief in cancer cases. It was tentatively determined that the period of this limited distribution should be at least one year, and the distribution procedure provided for the acquisition, by Dr. Eddy of the Public Health Service, of further clinical data from physicians who used it for the purpose stated. Dr. Eddy has made careful analysis and classification of the data received thus far, and consideration is being given to a plan which envisages somewhat more extended distribution of Metopon whereby it will be available in drug stores, on a physician's bona fide prescription, for the relief of chronic pain incident to other diseases besides cancer. It will be remembered, however, that entirely aside from any question of patent control, the production and distribution of Metopon is subject, as always, to the provisions of the Federal narcotic laws.

The second class of new drugs is that of the synthetic analgesics, produced without reference to opium or morphine as a parent substance. The first example of this class to be brought to our attention was Demerol. This drug was found to have habituating qualities comparable to morphine but, because it did not, even at the time it was released as a new drug by the Food and Drug Administration, come within the statutory definition of a narcotic drug, it was not at that time subject to the Federal narcotic laws. A special statute was therefore necessary and, under an act approved July 1, 1944, Demerol, under the coined term

"isonipecaïne," was made subject to the Federal narcotic laws to the same degree as is morphine, excepting that no provision was made for the use of Demerol in the so-called exempt (non-prescription) remedies.

The discovery, production, and distribution of Demerol foreshadowed the discovery and possible production for medical use of other synthetic analgesic drugs. Foreseeing this development, and considering the likelihood that other synthetic analgesics would have habituating qualities, the Treasury Department, upon the suggestion of the Ways and Means Committee of the House of Representatives, completed a proposed bill designed to cover under the Federal narcotic laws, by carefully regulated administrative procedure, each new drug as it was sought to be introduced into trade supply channels, provided the new drug was found to have habituating qualities similar to morphine or cocaine. It will be remembered that, about one year after the enactment of the special statute making the narcotic laws applicable to Demerol, the Department of Commerce made available to the American pharmaceutical industry the publication known as P. B. 981 which outlined processes and other information concerning such synthetic analgesics as amidon and Bemidon. Experimental production and clinical study of amidon was instituted shortly thereafter, and in the spring of 1946 the Public Health Service commenced its clinical studies of amidon.

Congress enacted the bill into law as Public Law 320—79th Congress, approved March 8, 1946. Section 1 of this act provided for the covering under the Federal narcotic laws of any drug found by the Secretary of the Treasury, after due notice and opportunity for public hearing, to have an addiction-forming or addiction-sustaining liability similar to morphine or cocaine, and proclaimed by the President to have been so found by the Secretary. The result of the Public Health Service tests of amidon, made available to the Bureau of Narcotics in April, 1947, indicated that the drug had the requisite habituating qualities and, following the statutory procedure, the Secretary made his finding accordingly. The President proclaimed such finding as of July, 31, 1947, the proclamation being published in the Federal Register as of August 2, 1947, and the Federal narcotic laws immediately became applicable to amidon. The same procedure is applicable to other synthetic analgesics now being, or which may be, studied, provided they are found to possess the requisite habituating qualities.

The procedure by which these new synthetic analgesic drugs with habituating qualities may be brought within the purview of the Federal narcotic laws with reasonable promptness is very important to that degree of control of domestic production and distribution which is designed to make these drugs available only for proper medical purposes within the United States. It is also desirable, and necessary, that the international traffic in such synthetic analgesics be likewise controlled pursuant to the system of international control currently applied pursuant to the narcotic drug conventions, notably the Geneva Convention

of 1925 and the Convention for Limiting the Manufacture and Regulating the Distribution of Narcotic Drugs of 1931. While these two Conventions provide machinery for extending their terms to new analgesic opium derivatives, such machinery is not made applicable to the synthetic analgesics. The Commission on Narcotic Drugs of the United Nations has recommended, in its report to the Economic and Social Council of that organization, that a draft Protocol be prepared and circulated to all Governments concerned, with a view to the said Protocol being brought into force at the earliest possible moment. The Protocol would be a separate international instrument to cover new drugs which do not fall under present Conventions, the procedure being somewhat analogous to that provided for covering new drugs under our narcotic laws, except that the ultimate fact-finding body is the World Health Organization. Every effort is being made to expedite action on this proposal in order that appropriate regulation of international movement, as well as control of world production, of the new synthetic analgesic drugs may be achieved as soon as possible, thus reducing the opportunity for diversion of such drugs to non-medical purposes.

A further consideration in the general plan of achieving complete control of the traffic in the dangerous synthetic analgesic drugs, is the question whether State narcotic laws can be made applicable to such drugs. It seems to be a rather slow process to achieve 100 per cent coverage of the narcotic enforcement problem in this respect, since even sixteen years after the model Uniform State Narcotic Act was drafted, there still remain some three or four states which have not adopted this legislation, or legislation which the Bureau of Narcotics deems of comparable efficacy. It is very doubtful that those states which have already adopted the Uniform Narcotic Law will accept a State administrative procedure similar to that in the Federal sphere, providing for a reasonably prompt coverage under their existing State law of the synthetic drugs found to be dangerous. It is probable that the new synthetic drugs will be covered, if at all, under these laws, one by one, as their dangerous habituating qualities become widely known, and it must be remembered that the State legislatures generally meet biennially and there is frequently a lack of interest in this important legislation. As far as I am informed, a movement looking toward the possible application of a State narcotic law to amideon has been started in only one state, New York.

In conclusion, I express the hope that any manufacturer or pharmaceutical laboratory which undertakes experimental study of a new synthetic analgesic will, subject to appropriate protection of its priority therein, submit complete data on the drug to Dr. Eddy for study and report by the Committee on Drug Addiction and Narcotics of the National Research Council. In addition, and even independently of a determination of the comparative analgesic efficacy of such a drug, it is most important to make a determination relative to the habituating qualities thereof, before the drug is distributed for medical use.

METHADONE IN INTERNAL MEDICINE

By RAYMOND N. BIETER AND STANTON A. HIRSH*

*Departments of Pharmacology and Internal Medicine,
University of Minnesota Medical School, Minneapolis, Minnesota*

THE release of methadone (10820) by the Food and Drug Administration about a year ago opened a new and interesting chapter on the action and uses of analgesics, particularly of the morphine type. Since then, other derivatives have been made available for experimental study. It is the hope that investigators will be able to discover the drugs of maximum use with the minimum untoward reactions, for routine or special uses. It appears at present that rather extensive clinical trials will be necessary to find the answers to these questions.

Among the available reports on the clinical usefulness of methadone in internal medicine are those of Scott, Kohlstaedt, and Chen,¹ Kirchhof and David,² Ishmael and Stacy,³ Bercel,⁴ and Troxil.⁵ All of these investigators found methadone to be a powerful analgesic. Milligram for milligram, it is probably more active than morphine. In most clinical investigations, a small number of patients have been studied in which both methadone and morphine or another opium alkaloid have been used. A considerable number of patients have shown a preference for methadone over morphine, meperidine and codeine. Reversed preferences have also been observed. Whether there exists a striking preference for any one of these drugs is still not proven. Likewise, some patients have been observed in whom one or more of the above analgesics have failed. In these instances, methadone has not always been the least active drug. And again, certain patients who have been unable to tolerate morphine have taken methadone and have shown satisfactory analgesia with no side effects. Therefore, methadone, as an analgesic, cannot be dismissed lightly.

In the overall picture presented by these investigators, complete relief of pain was obtained in 50 to 70 per cent of patients, moderate relief in an additional 20 to 30 per cent, and failures in from 3 to 20 per cent. If one selects the patients in "general medicine" to whom the drug was given from the reports of Scott *et al.*, and Troxil, one finds that moderate to complete relief was obtained in over 90 per cent patients. On the other hand, Batterman⁶ reported in 180 patients that control of pain by oral administration was produced in only 40 per cent of trials, and by parenteral injection in 76 per cent of trials. Inasmuch as no description of the type of patient was included and no data given as to whether control of pain was moderate and/or complete, these results cannot be adequately evaluated.

* Grateful appreciation is accorded to Abbott Laboratories, Ciba Pharmaceutical Products, Inc., Eli Lilly and Company, Lederle Laboratories, Inc., Parke, Davis and Company, G. D. Searle and Company, and Winthrop-Stearns, Inc., for research grants given to the Regents of the University of Minnesota for studies in clinical pharmacology.

The data recently reported by Troxil have been rearranged and the cases falling only in the field of general medicine have been selected for this purpose. The cases have been further grouped according to three dosage levels, 5, 7.5, and 10 mg., and, finally, into hypodermic and oral administration. This is shown in TABLE 1. In the hypodermic group, it

TABLE 1
ACTION OF METHADONE

<i>Method of administration and dose in mg.</i>	<i>Analgesia</i>			<i>Side effects</i>					
	<i>None</i>	<i>Fair</i>	<i>Good</i>	<i>Sedation</i>	<i>N&V</i>	<i>Dizzi- ness</i>	<i>Miosis</i>	<i>Euphoria</i>	<i>Itch- ing</i>
IN HOSPITAL PATIENTS:									
Hypo- dermic	5.0	7	15	59	6	1		1	
	7.5		5	34		2			
	10.0		4	30	2	1			1
Total of 154 patients; 14 side effects in 11 patients.									
IN AMBULATORY PATIENTS:									
Oral*	5.0	4	3	10	1	3	1		
	7.5			2	1		1	1	
	10.0	3		12	2	2	4	2	1
Total of 34 patients; 19 side effects in 13 patients.									

* As tablet, capsule, elixir.

can be seen that failures occurred only in the 5-mg. dose group. In the oral group, failures occurred also in the higher dose level. Although the total number of cases is small, it seems justifiable to state that failures occur more readily by the oral method of administration.

The lower incidence of adequate analgesia produced by the oral administration of methadone is probably not peculiar to this drug. The experience with oral meperidine on Dr. Cecil Watson's service at the University of Minnesota Hospital is that it is an unsatisfactory analgesic. The analgesic potency of morphine given orally in man is unknown. A theoretical explanation for the unsatisfactory analgesia produced by the oral administration of these drugs may be found in the study of Ercoli and Lewis.⁷ These investigators studied the analgesia produced by morphine, meperidine, and other analgesics in the rat, using a modified Hardy-Wolff apparatus. With meperidine, they found that it required an oral dose 3 to 5 times the size of the hypodermic dose to produce comparable analgesic effect, whereas with morphine the comparable oral analgesic dose was about 40 times the size of the hypodermic analgesic dose. One might, therefore, suspect that in man, too, morphine would be an unsatisfactory analgesic by mouth. With both morphine and meperidine given to rats orally, Ercoli and Lewis found no satisfactory relationship between dose and effect.

TABLE 1 also lists the side effects, which are, for the most part, mild. By hypodermic administration, 14 side effects were noted in 11 patients out of a total of 154. Orally, 19 side effects were produced in 13 patients

out of a total of 34. Thus it can be seen, as others have also observed, that side effects are more common following oral administration. Minimal side effects were also observed by Scott *et al.*, and by Kirchhof and David. It is to be noted that, in striking contrast with morphine, methadone produces sedation only occasionally. For this reason, this action of methadone has been grouped with the side effects.

Bercl, on the other hand, reported, in addition to side effects similar to the above, several reactions that were more severe. These were two cases of circulatory collapse which in his opinion were due to a sensitivity. Both of these patients recovered. A similar type of reaction, as possibly being due to methadone, was reported to the writers several months ago by Dr. Ben Sommers of St. Paul. He has consented to the following inclusion of this case.

Patient M. L., male, age 54. For the pains of *tabes dorsalis*, this patient was given hypodermic injections of methadone in doses of 10 mg. at 7 p.m., 20 mg. at 9 p.m. and 20 mg. at 3 a.m. The pain was relieved. At 10:30 the next morning the patient was nauseated, walked with a stumbling gait, was slightly cyanotic, but had and maintained a normal respiratory rate. He gradually became unconscious. At 5 p.m. he was admitted to a hospital. At that time, his pulse was 80 and his blood pressure 40/0. Despite vigorous intravenous therapy, the patient made no improvement and died 24 hours later. Diagnosis: circulatory shock and cerebral irritation.

Autopsy findings: Gross cerebral edema; petechia in pericardium; aspiration pneumonia.

Shortly thereafter, the writers were only mildly surprised to learn of a similar type of circulatory shock which followed a dose of morphine. The following summary of this case was given to the writers by Dr. C. N. McCloud of St. Paul for inclusion in this manuscript.

Patient I. B., female, age 59, 170 lbs. At 5 a.m. for an acute attack of nausea and vomiting with pain, patient was given 15 mg. of morphine sulfate by intravenous injection. Before the injection bl. pr. was 170/80 mm. Hg, pulse 108 and respiratory rate above normal (overbreathing). Patient quickly obtained pain relief and sedation from the morphine. Respiration appeared normal. In the afternoon, patient was found comatose, with blood pressure 60/0, pulse rate 84, respiration as in normal sleep. Thirty minutes later, in a hospital, the blood pressure was 30/0 with a pulse of 90 which was thready. Respiration was like that in normal sleep. The Babinski was positive. Patient began to become restless so 10 mg. of morphine was given by subcutaneous injection. Extensive intravenous fluid therapy was instituted. During the next few hours, the blood pressure gradually rose to 90 mm. Hg. During the next two days the blood pressure gradually returned to 160 mm. Hg systolic, and consciousness also slowly returned. The neurological diagnosis was cerebral edema.

Similar circulatory reactions with meperidine have also been observed in the Twin City area of Minnesota, especially in obstetrics. A mild circulatory collapse following an injection of meperidine occurred on Dr. Watson's service in the University Hospital recently. He has consented to our inclusion of the following summary:

Patient M. L., female, age 35, 193 lbs. On admission bl. pr. 124/78 mm. Hg, pulse 72 per minute. 3 p.m. Meperidine 100 mg. hypo given for pain. 3:40 p.m. patient complained of feeling queer and that she could not open her mouth. Pulse became thready, rate 48 per minute, bl. pr. 90/85. Respiratory rate 16. Skin cold and clammy, lips cyanotic. 4 p.m.

0.5 cc. epinephrine HCl hypodermic injection. 500 cc. glucose and saline by vein. 4:45 p.m. bl. pr. 110/58, pulse 92. 5 p.m. patient feels much better, bl. pr. 102/58, pulse 92. 9 p.m. blood pressure 106/64.

Subsequent recovery was uneventful.

Thus, it appears that evidence of circulatory collapse can be seen following the administration of these three analgesics to man. On second thought, this is not too surprising. In animals, all of these drugs can be shown to produce striking blood pressure falls from which recovery may be rather slow. Data on the mechanisms of action of these drugs point in a similar direction. Krueger, Eddy, and Sumwalt⁸ in 1940 stated in their monograph on the opium alkaloids that in some instances morphine appears to increase parasympathetic activity, or, as other investigators have suggested, to inhibit sympathetic activity. Bernheim and Bernheim⁹ in 1936 reported that morphine inhibits cholinesterase in the brain of animals. It is not unreasonable to suppose that, in some instances, morphine may increase parasympathetic activity much more than in most others and, thus, hypotension could result at times. This action may be at least part of the answer to the action of morphine on the bronchi, gastro-intestinal tract, and urinary bladder. Constipation has also been noted as a side effect in a few patients following methadone. Meperidine is commonly thought of as a smooth-muscle depressant. The fact that it has a hypotensive effect is in accord with this viewpoint.

On the other hand, all investigators have not found meperidine to relax all smooth muscle structures, even in experimental animals. In man, a striking spasm-producing effect, or the lack of an antispasmodic effect, was recently reported by Gaensler, McGowan, and Henderson.¹⁰ Following surgery with T-tube drainage in the biliary tract of man, they found that 100 mg. doses of meperidine produced almost as great an increase in the pressure within the system as 10 mg. doses of morphine. Possibly, again, meperidine can at times be more like morphine in its action than we ordinarily think of it at present. Scott, Kohlstaedt, and Chen found, in animals, that the effect of methadone on the circulation, in its entirety, was more like morphine than like meperidine. Finally, Shideman and Johnson,¹¹ studying all three drugs in the dog, found all to be hypotensive, although with repeated injections the responses show differences. Evidence is insufficient to indicate whether one of these analgesics in man is more prone to produce hypotension than the others. On the basis of the frequency with which hypotension has occurred within the writers' knowledge, the drug most apt to produce this effect is meperidine.

Another serious untoward reaction that is bound to occur with methadone is respiratory depression. Such a case has been observed by Dr. Harvey O'Phalen of Minneapolis, Minnesota, with whose permission the following abstract is included.

Patient J. R., age 15. Amputation of the leg following a compound fracture. For two months postoperatively, the following analgesics were used in acceptable doses 3 to 5 times

per day: morphine, codeine, Pantopon, meperidine, and various combinations of these with the coal tar analgesics. Upon the suggestion of the writers, methadone was substituted. In 2.5 mg. hypodermic doses, adequate analgesia was produced with no side effects. Four to five doses per 24 hours were satisfactory. An occasional 5 mg. dose was tried. Doses of this size appeared to be unnecessary. One evening, by mistake, the patient was given a 25 mg. hypodermic injection. Within an hour, his respiratory rate had decreased from 20 to 11 per minute. Several large cups of strong black coffee were given to the patient. Within another hour, his respiratory rate rose to 18 per minute and he was less comatose. He was then left alone, went to sleep and slept well throughout the night. His respiratory rate did not drop below 18 per minute thereafter. Thirty hours later, the patient asked for another analgesic hypodermic and was put back on his usual schedule with methadone. Whether this patient had acquired some tolerance to these drugs during 2½ months of analgesic therapy or whether he was resistant to methadone is not known. Everyone is familiar with these effects produced by morphine, especially in overdose. Meperidine, on the other hand, in the acceptable dosage range, seems to be less depressant to respiration. From this standpoint, it appears to be the safest of the three.

Several other types of effects and untoward reactions have occasionally been observed in the University of Minnesota Hospitals, which are not included in the above summaries. These effects have occurred, chiefly, following the administration of several doses. Similar reactions have been reported by other investigators. Examples are given below.

(1) Patient G. L., female, age 19. Diagnosis: abscess of right incisor tooth, with severe pain. 9 p.m., methadone 10 mg.—no analgesia. 1 a.m., methadone 10 mg.—good relief for 5-6 hours.

(2) Patient I. H., female, age 28. Diagnosis: thrombophlebitis, with severe pain. Methadone 5 mg. hypodermic injections at 4 to 5 hour intervals. The first two doses produced no analgesia. After the third dose, analgesia was good. Coincident with pain relief, drowsiness occurred which persisted for more than 24 hours and then disappeared. Good analgesia with no side effects occurred for nine additional days. Drug was then discontinued. Following meperidine, 150 mg. hypodermic injections, this patient obtained much less analgesia, but more hypnosis and also nausea; with morphine 10 mg., the analgesia was also inferior and mild side effects were common.

(3) Patient J. M., male, age 55. Diagnosis: hemiplegia with thalamic pain. Methadone 2.5 mg. capsules at 1 hour intervals for 1 day, then 5 mg. capsules at 4 hour intervals for another day. No analgesia produced. Sedation very good. Patient gradually developed slurring of speech which progressed almost to a complete aphasia. Meperidine, 150 mg. hypodermic injection, no analgesia. Morphine sulfate 15 mg. hypodermic injection gave analgesia. Untoward effects were Jacksonian seizures.

(4) Patient F. U., male, age 67. Diagnosis: painless jaundice. At 5 p.m. patient was given a 10 mg. hypodermic injection of levomethadone in preparation for peritoneoscopy. One-half hour later patient developed euphoria and disorientation, and could not speak easily. Pupils were constricted. Slept well all night. The next morning, he still showed euphoria. No other effects noted.

These cases necessitate a consideration of mechanism of action and comparison with other analgesics for adequate appraisal. The result obtained in Case 1 has been repeated not infrequently. This effect could be due to lack of absorption following the first dose or to a certain cumulation, which might be necessary in this particular individual in order to obtain the desired action. In other words, a certain level or concentration of drug in the area of the central nervous system essential for perception of pain may be necessary. This type of reaction (no analgesia fol-

lowing the first dose, good analgesia following the second and subsequent doses) has been observed repeatedly with hypodermic injections of morphine. Under these circumstances, it would appear impossible to indict either drug. A similar explanation might be justified in interpreting Case 2. In this case, the increase in action with repeated doses also involved the production of a temporary sedation. This case also illustrates the comparative effects of three analgesics. In patient 3, comparative analgesic effects are again shown. This type of response to methadone, *i.e.*, severe nervous system side effects, undoubtedly is to be associated with the degree of brain injury due to the disease. A similar explanation may be used to explain the lack of respiratory depression of methadone, compared to that of other narcotics and barbiturates in bulbar and spinal poliomyelitis, as was reported by Troxil. Altogether too little is known of the relative depressant effects of barbiturates and the morphine type of analgesics in patients with brain injury. Patient 4 illustrates a type of untoward reaction seen not uncommonly in patients with liver disease. Dr. Cecil Watson has, likewise, seen severe untoward reactions with morphine. As a result, it is his opinion that morphine is contra-indicated in acute liver disease of this type. This patient was given levomethadone with the hope that it would be tolerated more satisfactorily than morphine. That appears not to be the case. Whether these reactions to morphine and methadone are associated with the lack of ability to detoxify or change these drugs in certain liver diseases is unknown. Certainly, it should be further investigated. From the standpoint of this group of untoward effects and the hoped-for analgesia, it would seem that there is no fundamental difference between any of the three analgesics, morphine, meperidine, and methadone. Considerable work, both experimental and clinical, is necessary to determine whether one is safer or another more toxic in patients who have serious pathology in one or more vital organs.

A study of methadone derivatives has also been instituted. Because the possible dosage level of these derivatives is very much alike, or identical, these drugs have been taken over to the University Hospital labeled Methadone A, Methadone B, Methadone C, etc. The clinical staff members order these drugs and one of us (Dr. Hirsh) observes the patients. Racemic methadone has also been included. In this way, a minimum amount of bias will creep into the determination of indications and interpretation of side effects. The data obtained to date are decidedly a preliminary report and should be observed from that standpoint. TABLE 2 shows a summary for each drug. The number of cases is insufficient to divide them according to 5, 7.5, and 10 mg. doses. All of the patients were hospitalized and all received the analgesics by hypodermic injection. It is to be noted that—with a limited number of cases—there appears to be no striking difference to date between methadone, levomethadone, and levo-*iso*-methadone. The drug, K4710V (10720), does, however, appear to differ from the above three. Sedation with this

TABLE 2
ACTION OF METHADONE AND DERIVATIVES

Drug, hypodermic administration, 5, 7.5, and 10 mg.	Analgesia			Sedation		Side Effects	
	None	Fair	Good	None	Good	N&V	Dizziness
Methadone*	2	1	13	15	1	2	1 (in 3 patients)
l-Methadone†		2	24	23	3	3	(in 6 patients)
l-iso-Methadone‡	2	3	12	13	4	1	1 (in 6 patients)
K4710V§ (10720)		2	22	8	16	3	4 (in 4 patients)

The drugs were generously supplied by the following pharmaceutical concerns: * Eli Lilly and Company; Winthrop-Stearns Inc. † Abbott Laboratories; Winthrop-Stearns Inc. ‡ Winthrop-Stearns Inc. § Winthrop-Stearns Inc.

drug is far more common and, as a result, this effect has been taken out of the "side effect" group and placed beside analgesia. In this respect, this drug is more like morphine and meperidine than the others. It should be noted that patients who want sedation as well as analgesia prefer this drug to the members of the methadone group. Not all patients, however, have this preference for sedation. In this group of patients, it was noted that two patients out of a total of seventeen receiving levo-iso-methadone obtained no analgesia until after the second dose had been administered. Then all of them reported that the analgesia was better than that from any other previously used drug.

Again it is emphasized that these data on the methadone derivatives are preliminary. It is hoped to continue this type of study for a year or more.

Summary

1. Methadone has been shown by a number of investigators to be a potent analgesic for the control of pain in internal medicine.
2. Whether its analgesic action is superior or inferior to that of morphine and/or meperidine remains to be proven.
3. From the standpoint of preference on the part of some patients and its ability to produce analgesia when other drugs fail in other patients, it has earned a place in human therapy.
4. Unlike morphine and meperidine, it is devoid of consistent hypnotic effects.
5. Untoward reactions have been observed following its use by all investigators in clinical studies. These reactions are again like those following morphine and meperidine.
6. The relative danger of these three analgesics from the standpoint of untoward reactions is unknown.
7. Instances of possible severe untoward reactions in man following the administration of each of the three analgesics have been reported.
8. Levomethadone, levo-iso-methadone, and K4710V (10720) have also been shown to be good analgesics in man.

References

1. SCOTT, C. C., K. G. KOHLSTAEDT, & K. K. CHEN. 1947. Comparison of the pharmacologic properties of some new analgesic substances. *Anesth. & Anal.* 26: 12.
2. KIRCHHOF, A. C., & N. A. DAVID. 1947. Clinical trial of a new synthetic heptanone analgesic (Dolophine). *West. J. Surg. Obst. & Gyn.* 55: 183.
3. ISHMAEL, W. K., & J. R. STACY. 1947. The use of Dolophine in the control of pain in bone and joint diseases. *J. Oklahoma Med. Assoc.* 40: 454.
4. BERCEL, N. A. 1948. Clinical trial of 10820, a new synthetic analgesic. *Dis. Nerv. Syst.* 9: 15.
5. TROXIL, E. B. 1948. Clinical evaluation of the analgesic, methadon. *J. A. M. A.* 136: 920.
6. BATTERMAN, R. C., & A. M. OSHLAG. 1948. The effectiveness and toxicity of methadon. *Fed. Proc.* 7: 206.
7. ERCOLI, N., & M. N. LEWIS. 1945. The time-action curves of morphine, codeine, dilaudid and Demerol by various methods of administration. *J. Pharm. & Exp. Therap.* 84: 301.
8. KRUEGER, H., N. B. EDDY, & M. SUMWALT. 1941. The pharmacology of the opium alkaloids. *U. S. Pub. Health Rept., Suppl.* 165 (1): 69.
9. BERNHEIM, F., & M. L. C. BERNHEIM. 1936. Action of drugs on the choline esterase of the brain. *J. Pharm. & Exp. Therap.* 57: 427.
10. GAENSLER, E. A., J. M. MCGOWAN, & F. F. HENDERSON. 1948. A comparative study of the action of Demerol and opium alkaloids in relation to biliary spasm. *Surgery* 23: 211.
11. SHIDEMAN, F. E., & H. T. JOHNSON. 1948. Acute vascular tolerance to morphine, isonipecaine (Demerol) and methadon (amidone) in the dog. *J. Pharm. & Exp. Therap.* 92: 414.

USE OF NEW SYNTHETIC ANALGESICS IN SURGERY

By E. A. ROVENSTINE

*Divisions of Anesthesia, New York University College
of Medicine and Bellevue Hospital, New York, N. Y.*

ANALGESICS have a role in surgery that encompasses more than reducing the sensibility to pain. They serve two important purposes. This duality is expressed in their utilization to provide safer, less disagreeable anesthesia. They are effective also in alleviating discomfort before and after surgical manipulations. In the former instance, analgesics are among the rather extensive group of drugs that are employed for pre-anesthetic medication.

Pre-anesthetic medication is a descriptive term indicating drugs used in the period immediately before an anesthetic agent is administered. In general usage, it identifies the drug, or drugs, given hypodermically shortly before the patient leaves his hospital bed for the operating room. It is almost a tradition in the majority of hospitals that this "hypo" contain morphine and a belladonna derivative, atropine or scopolamine. A broader interpretation of the term, pre-anesthetic medication, includes a wider list of drugs and many diverse purposes. It is the preparation of the patient for anesthesia and surgery. This may be a hypnotic to insure a restful sleep the night before operation, it may mean insulin and glucose to fortify the diabetic, quinadine for the patient with heart disease, or any of a number of drugs that mitigate against anticipated complication during anesthesia and surgery. Pre-anesthetic medication is given in an attempt to make "the patient safe for anesthesia." Its importance should not be minimized or subjugated to attempts to make "anesthesia safe for the patient."

For the purpose of this discussion, the two most needed effects from pre-anesthetic medication will be considered. These are, to reduce the amount of anesthetic agent required to produce the optimal degree of narcosis for the proposed operation, and to protect the patient from mental discomfort in the immediate pre-operative period. Among the new synthetic analgesics, only Demerol has been found to be especially useful as a pre-anesthetic drug for these purposes. Methadone appears to be without the euphoric and sedative action necessary to alleviate pre-operative apprehension.

Demerol has been tested widely as a substitute for the opiates given in the pre-anesthetic period. Such substitution has long been desired since the opiates have many untoward reactions interfering with the anesthetic regime and convalescence. It has not followed that Demerol meets all the requirements as a pre-anesthetic drug. In the doses considered safe, many active young adult patients will not reach the surgi-

cal amphitheater in a fearless, confident psychic state. With the opiates given properly, this is less frequently the case. The most favorable patients for Demerol premedication are the elderly and those for whom the opiates are withheld. This group is steadily increasing in numbers and includes patients who might suffer severe consequence from respiratory or circulatory depression, urinary dysfunction, or profound hypnosis.

It has been the clinical observation that Demerol serves to decrease the amount of drug needed for anesthesia. Although this observation has not been confirmed by controlled experiment in man, it was found in our laboratories that it was true for the dog. The amount of cyclopropane in arterial blood was reduced from one-third to one-half that required for a given plane of anesthesia after the dog had received 10 mg./kg. of Demerol. Similarly, the quantity of ether needed was decreased some 25 per cent, and the anesthesia time after a given dose of intravenous pentothal was prolonged more than 100 per cent. These results with Demerol compare favorably with those observed when the opiates are used. This despite the fact that morphine is the drug par excellence for reducing metabolic activity. It has been the contention, since 1869, when the French physiologist Claude Bernard demonstrated the value of sedative drugs for pre-anesthetic medication, that less anesthetic agent was required as general metabolism was progressively depressed and reflex irritability further obtunded. Since Demerol is inferior to morphine in reducing metabolism, other factors must have an important bearing on the results.

When pre-operative pain is present, it is usually controlled by Demerol in appropriate doses. The analgesic effect, however, is not so constant and of shorter duration than may be had with morphine in the amounts regularly given.

When Demerol is given parenterally in doses of 100 mg. approximately one hour before induction of anesthesia, about one-half of the patients in good physical condition in the 20 to 60 year age group will reach the operating room calm and without evidence of emotional disturbance, but with an alertness permitting response to ordinary conversational questions. This represents the optimal effect from such medication. The percentage of satisfactory responses to the same amount of drug increases to some 75 per cent if scopolamine is combined with it. In the older patients, more than 50 mg. is seldom advised because of the danger from respiratory and circulatory depression. The drug is somewhat more efficacious in the old-age group and is usually used without adding a belladonna derivative. The use of Demerol orally has not given good results for pre-anesthetic medication.

The second role of the analgesics in surgery is concerned with the control of postoperative pain. There is no chapter in surgical history which tells a less inspiring story than the one on the progress in making the patient comfortable after operation. The practice of writing orders for the

inevitable "MS $\frac{1}{4}$ q 4 hrs. or PRN"* prevailed through many surgical years. Such orders were usually written in the operating room before the patient had fully recovered from anesthesia or reached his bed. Recently, there has been a scientific effort to eliminate this empirical routine. These efforts have resulted from the tardy realization by those entrusted with postoperative care that sedatives and analgesics interfere with convalescence, and increase nursing attention and hospitalization. Moreover, their role in surgical morbidity and mortality may be of considerable magnitude. A better understanding of pain mechanisms, and the introduction of new drugs and procedures to alleviate pain and discomfort, have increased recently the efforts to keep patients comfortable without large amounts of depressing drugs. The value of permitting the patient to utilize his natural defenses against impending complications is recognized. Changes in position, breathing exercises, coughing, frequent exercises, and even early ambulation have served to reduce morbidity and mortality. A more careful use of sedative and analgesic drugs has been the result. Postoperative pain therapy has emerged upon a more rational basis. The early empiricism is being discarded rapidly and a sound scientific rationale is being adopted.

Demerol has been used rather extensively in Bellevue Hospital as a postoperative analgesic. When the need for such medication is determined, and it is far from routine, the parenteral use of Demerol is suggested. It has the advantage over morphine of rarely producing respiratory depression, of being antispasmodic, and interfering little with the cough reflex. It is given in doses of 50 to 100 mg., and in such amounts may be effective for 2 to 3 hours. When the drug is given orally, it is much less effective and is rarely chosen over other mild analgesics. The side effects from Demerol include vertigo, nausea, emesis, and syncope, but with rather a low incidence compared with morphine. Circulatory depression and a decrease in respiratory rate may follow larger doses and the drug is used more cautiously for aged individuals. It is not often given to children.

The effectiveness of the synthetic analgesic methadone on postoperative pain has been investigated briefly by Batterman and his associates in Bellevue Hospital. The drug was administered in one dose of 2.5 to 20.0 mg., or similar doses, repeated several times daily. More than 2,000 parenteral and as many oral doses were given. It was the impression that methadone is a useful drug for controlling pain following surgery. It is not recommended for prolonged use in elderly patients, since they may more often develop untoward reactions such as anorexia, nausea, vomiting and dizziness, and with continued use more serious mental confusion and toxic psychoses have been observed. Methadone seems most effective in 10-mg. doses given hypodermically. Severe postoperative pain is readily controlled in a high percentage of patients with this amount. The side effects from methadone are minimal, and its sedative and eu-

* Morphine $\frac{1}{4}$ grain every 4 hours, or as needed.

phoric action is less marked than from other potent analgesics. When given orally, the elixir proved most useful and the favored dose was 5 mg. The drug is usually effective for three or more hours, depending on the severity of pain. The experience accumulated to date is not comprehensive and does not warrant an effort to assign a definite role for methadone in controlling postoperative pain. It is altogether likely that it will not be a substitute for morphine since it is not yet proven superior to it. Like Demerol, it may find its place in selected types of pain following surgery where it can be carefully individualized for the patient.

Tridione, now used for its anticonvulsant action, was first investigated as an analgesic. Its analgesic properties are well established, although it is not to be compared to the new synthetic drugs already discussed, or with the opiates. Its effective action is less predictable. The drug, given intramuscularly in doses of 10 cc. of 10 per cent solution, is practically free of side effects other than sedation. It is not useful when pain is severe and its mild analgesic action from the dose cited is scarcely of more than 2 hours' duration. It has none of the euphoric properties characteristic of morphine analgesia. Its usefulness as an analgesic when given orally for postoperative pain is slight. Its limitations and indications are as yet undetermined.

This discussion would not be complete without mention of intravenous procaine. Procaine is not a new drug, but it is only recently that it has been widely used for analgesia by intravenous administration. The enthusiasm that has accompanied this use has its basis in clinical observations. To date, there is no agreement as to how its analgesic action is accomplished. The site of action is not determined experimentally. It is obvious that the more spectacular results have been with the control of pain associated with trauma. Patients with burns, fractures, sprains, and those with pain at the site of surgical manipulations are more regularly benefited from intravenous procaine. This observation suggested that analgesia follows the extravasation of procaine through damaged capillaries where it reaches nerve endings in the perivascular areas at the site of injury. The rapid hydrolysis of procaine is contrary to this concept. It has been suggested, also, that the central effect produced by procaine was responsible for the analgesia. It has been determined by the Hardy-Wolff-Goodell technique that the pain threshold is elevated in patients having received large subcutaneous doses of procaine, an elevation more prolonged than the duration of local tissue analgesia. Using a similar technique, it was determined in our laboratory that one gram of procaine given intravenously, as used clinically to control pain, failed to raise the pain threshold as much as a therapeutic dose of acetyl-salicylic acid. This is not in keeping with the analgesic action observed and throws doubt on the importance of the pain threshold as a modality for evaluating analgesic drugs, or points to another mechanism to explain the results with procaine.

Since the whole problem seemed confused, an investigation was un-

dertaken in our laboratories and clinics to clarify the pharmacology of procaine and the products of its hydrolysis, diethylamino-ethanol, and para-aminobenzoic acid. Brodie and his associates devised chemical methods for identification of these substances. With these procedures it was demonstrated that, in man, procaine is rapidly hydrolyzed to the alcohol and acid after intravenous injection. They learned further that urinary excretion of injected procaine is negligible, but that 75 to 95 per cent of the predicted para-aminobenzoic acid injected in procaine is excreted unaltered in the urine. This was in contrast to the 20 to 35 per cent of the predicted amount of diethylamino-ethanol which could be isolated from urine. It was obvious, then, that the diethylamino-ethanol product of the hydrolysis of procaine is further metabolized *in vivo* in a manner not yet determined. This, and the fact that procaine persists in plasma for a much shorter time than diethylamino-ethanol, suggests that the latter drug may be the pharmacologically active agent rather than the parent drug.

It was logical, then, to prepare diethylamino-ethanol for oral, intramuscular, and intravenous use, and study its toxicological and pharmacological action. Emphasis was given to experiments leading to an evaluation of its possible analgesic effects. Such a study is now in progress. Initially, it is being used to control postoperative pain.

Diethylamino-ethanol produces no obvious toxic effects in man even when large doses are given. This is in sharp contrast to procaine, where the most serious criticism of its use intravenously is the frequency of toxic reactions. Also, there are some definite indications that diethylamino-ethanol is effective as an analgesic when given orally or intramuscularly.

The difficulties of evaluating analgesia in the postoperative period are readily appreciated, and it is realized that many observations need to be completed to reach conclusions. The experiments with diethylamino-ethanol completed here are still in the preliminary observation stage. The results merely suggest methods of further study and lead to early impressions.

After completing toxicology studies in animals and man (to be reported elsewhere) and determining doses probably effective, the initial studies on its analgesic properties were instituted. These included the administration of large amounts of the drug, 4 to 5 grams, in an infusion during surgery. The patient, surgeon or nursing staff were not informed that the drug had been given. Patients undergoing operative procedures that are regularly accompanied by postoperative pain of a severity requiring potent analgesics were selected. Morphine and other analgesic drugs were withheld postoperatively unless there was a complaint of pain by the patient. Other patients submitting to similar operations were given saline infusions only, and some were given 1 or 2 grams of procaine intravenously. No toxic symptoms have been noted following the administration of 4 grams of diethylamino-ethanol during this

study. The analgesic effects appear to surpass those observed with procaine. Two-thirds of patients so treated need no analgesic drug during the 24 hours following surgery. Because of the limited supply of the drug as yet available, continuous treatment has not been studied carefully. There are, however, several isolated instances where diethylamino-ethanol was given intravenously to unanesthetized patients in a single dose. A favorable analgesic effect was noted in a high percentage of patients and suggested a constant favorable response. In any event, the results obtained with diethylamino-ethanol, especially because of its relative absence of toxicity, make it mandatory to complete further studies.

USE OF THE NEW SYNTHETIC ANALGESICS IN OBSTETRICS

By BERT B. HERSHENSON

The Boston Lying-in Hospital, Boston, Massachusetts

THE purposes of this paper are three-fold: (1) to consider some of the essential principles basic to the clinical practice of obstetrical amnesia, analgesia, and anesthesia; (2) to present some experiences that we, at the Boston Lying-in Hospital, have had in trying to solve these problems; and finally (3) to record a brief report on our current use of Demerol.

Essential Principles Fundamental to Good Practice¹

The first and foremost principle is to give consideration to the welfare and *safety* of both mother and baby. Experimentation with new drugs on humans is a dangerous procedure and should not be undertaken in obstetrics unless the clinical team includes an anesthesiologist. I have no bone of contention with those humanitarians who seek to discover safer methods of providing pain-relief for the woman in labor. However, the emphasis should properly be placed on the knowledge, experience, skill, judgment, and attention of the members of the obstetrical team, rather than on the agent or technique employed.

The problem of fear and apprehension on the part of the prospective mother entering the hospital labor room should be met early. The psychic component is an important factor in intensifying pain-perception, as well as increasing pain-reactivity. The character of the environment, as well as the personalities of the obstetrician, anesthesiologist, and their co-workers, vitally influences this psychic component. A good principle to apply in the labor room is to put into practice some of the freedoms, *i.e.*, freedom from fear, freedom from apprehension, freedom from noise and careless chatter. This practice will go a long way toward producing the desirable pharmacological effects in accomplishing freedom from pain. Therefore, an essential principle is to effect maternal psychic sedation and amnesia, for the period of labor and delivery, consistent with safe clinical practice.

The pharmacological agents selected should permit the mother to rest quietly and peacefully between her uterine contractions and thus conserve her functional reserves. The amnesic and analgesic drugs selected should effect obtundation of the sensory phenomenon of pain. Some objective evidences of such therapeutic effectiveness are quiet reaction patterns, peaceful facial expressions of each mother, along with maintenance of a quiet, efficiently functioning labor room. The other specific sensory effects are relatively inaccessible to quantitative anal-

ysis. From a broad aspect, the analgesics are compounds which exhibit no similarity in chemical structure to each other, have nothing in common with regard to pharmacological action, and produce their effects through actions on different structures.

On the physiological side, we are dealing with two individuals, simultaneously, in whom reflex irritability is practically at opposite poles. On the one hand is the mother with increased reflex irritability, and on the other hand, the unborn child with reflex irritability at its lowest during the life-span. In addition, there are modifying factors which influence the patient's tolerance to the analgesics, *e.g.*, variations in sthenicity, nutritional state, efficiency of detoxifying mechanisms and organs of elimination, presence of increased thyroid activity, and others. Also, in the case of the unborn, there are factors of another order operating, as a state of immaturity or post-maturity, interferences with its life-line *in utero*, and others.

The objective of obstetrical amnesia and analgesia is not the attainment of complete comfort for all mothers. If this were so, then some mothers would have to be sacrificed due to extremes of drug depression. The complete individualization of selection of drug, dosage, and route of administration is the only path to safety. The state and duration of amnesia and analgesia sought for in the obstetrical patient must be differentiated from the degree of analgesia desired for the medical or surgical patient. In the case of the obstetrical patient, the level of drug depression reached is usually below the first stage where responsibility for drug administration must be assumed, not by the patient, but by the members of the obstetrical team.

The techniques employed for administration of the amnesic and analgesic agents should permit of controllability, yet be rapidly and pleasantly effective. The duration of desired clinical effect should be sufficiently predictable to offer the mother a pleasant induction and emergence. During and subsequent to maintenance, the amnesics and analgesics employed should be devoid of discomfort or hazards to both mother and baby.

The drugs, including the newer synthetics, should be non-toxic to both mother and baby. These agents should have received careful study in the experimental field on many and varied species of animals to determine lack of toxicity effects, before being applied to medical and surgical cases. Furthermore, only such drugs proven free of toxicity and serious undesirable effects experimentally should be exposed to extended clinical experience in the non-obstetrical clinical fields. Only then should these drugs be exposed to clinical obstetrical investigation in various dosage relationships, and by various routes of administration. In any case, we must preclude the acceptance of amnesics and analgesics which involve dangers of toxic, irritating, or undesirable effects disproportionate to their therapeutic value.

The observations reported from clinical obstetrical investigation must

include sufficient details to permit verification. Mere multiplication of inaccurate observations does not in any way render them accurate. The credibility of the data and the justification of the deductions are necessarily influenced by the experience of the investigators, their reputation as to disinterestedness, technical ability, and critical judgment.

The drug, or drug combination employed should be safe enough to permit the obstetrician to practice a conservative policy in the conduct of labor. Complicating symptoms, incident to the use of amnesic and analgesic drugs, during the *intra-partum* and *post-partum* periods, should be few and preferably absent. The modification of reactivity of the mother by the use of analgesics must not be so great as to sacrifice sane and safe obstetrical practice.

The newer synthetic analgesics should meet the requirements of the anesthesiologist. He must be able to meet the needs of mother, baby, obstetrician, and labor-room personnel. The newer synthetics should possess a reasonably wide margin of safety in providing satisfactory amnesia and analgesia, regardless of the gestational age or age of viability of the fetus. It is being realized that heavy basal narcosis for the mother may be detrimental to her unborn child. The drugs, including the newer synthetics, should be able to reduce reflex irritability of the mother without essentially diminishing fetal reflex irritability. The newer synthetic analgesics should not more deeply depress the vital functions of respiration and circulation, maternal and fetal, than the desired clinical state of reflex obtundation sought for. The minute-volume pulmonary ventilation should not be decreased as a concomitant effect in the use of drugs for pain-relief in labor. Depression of this order necessarily leads to hypoxia. Too much emphasis cannot be placed on the fact that the most potent depressant of cell activity is the induction of oxygen-want. Furthermore, not only central but also peripheral reflexes that play an important part in respiratory control should not be depressed because of the use of synthetic analgesics. Increased fetal movements *in utero*, marked change in fetal heart rate, or the appearance of meconium in the maternal birth-canal should never be caused by the use of analgesic agents for maternal comfort. Should the analgesic agent be established as a cause for such depression of fetal vital functions, then that agent had best be avoided in obstetrical practice. Ensuing complications of this order may well lead to increased obstetrical operative frequency, as well as other forms of avoidable meddling obstetrical practices.

The newer synthetic analgesics, ideally, should be retained on the maternal side of the tenable utero-placental barrier. These agents should not cause imbalance of reflex mechanisms associated with the autonomic nervous system. However, they should selectively depress the serous salivary and mucus secretory functions of the respiratory tract without altering its ciliary function. They should not induce nausea or emesis, nor cause depression of the cough reflex. The drugs should not

alter uterine functions. They should be devoid of inducing spasm of smooth muscle, *e.g.*, they should not cause prolonged spasticity of uterine muscle. This action not only potentiates maternal pain but constitutes a threat to the life of the fetus.

Periodic, intermittent use of the synthetics during the period of labor, *i.e.*, for approximately 24 hours, should not result in cumulation, should be free of tolerance-effect, and should not lead to addiction. These agents must be sufficiently predictable in safe effects so that administration is possible to all patients of the child-bearing period. This includes such factors as variations of gestational age and the presence of complicating disease-processes without fear of eliciting or potentiating untoward responses.

The newer synthetic analgesics and amnesics should permit exact control of the selected dose and administration by any desired route. When administered parenterally, reasonably satisfactory obstetrical amnesia and analgesia should be predictable in at least 90 per cent of the cases. This must be accomplished without obtunding the natural forces and protective mechanisms of labor.

It is desirable that these agents should possess the capacity of preventing toxic reactions when solid analgesic regional agents are used to complete the delivery. If inhalational agents are used for the latter purpose, these newer synthetic agents should permit controllable maintenance of the desired depth of narcosis without increasing the hazard of neonatal apnea.

The drugs selected for obstetrical amnesia and analgesia should be stable when in solution, and remain unchanged both chemically and physically during storage. The solutions of such drugs should be non-irritating when administered by injection, and free from allergic responses. They should be comparatively inexpensive and readily available. The aqueous solution of the drug should have constancy of action and predictability of reproducible effects. The newer synthetics should possess greater analgesic potency, amnesic effectiveness, but less toxic effects and fewer undesirable side-reactions with a greater margin of safety than other known available agents.

It becomes clear that the ideal agent, or combination of agents, is not available to meet all the requirements necessary for safe, controllable, and effective obstetrical amnesia and analgesia. There is a definite need for the development of new superior agents for the field of obstetrical amnesia and analgesia. The laboratory findings on several species of these newer synthetics must be known, and such data will be very helpful. But the evaluation of these newer agents by several experienced observers in the non-obstetrical clinical fields is even more essential. However, it must be emphasized that safety in this field of therapeutics rests on the knowledge, experience, judgment, and attention of the members of the obstetrical team rather than on the drug or combination of agents.

*Experiences in Solving These Problems at the
Boston Lying-in Hospital*

Our experience at the Boston Lying-in Hospital in trying to provide for safe pain-relief for the obstetrical patient goes back over a century—as a matter of fact, to the year 1847. Our practice is limited to the use of inpatients of the hospital, and not in domiciliary practice, nor for the ambulatory patient seen in the *ante-partum* clinics of the hospital. It is worth noting that the first obstetrical operation under anesthesia in America was performed by Doctor Walter Channing on May 5, 1847, at the Boston Lying-in Hospital. Dr. Channing's book,² *A Treatise on Etherization in Childbirth*, published in 1848, is now a classic. As he said, "It treats of a noble subject, the remedy of pain." Today we are demonstrating that this is still a subject deserving of our careful consideration.

We, at the Boston Lying-in Hospital, have been through various developmental phases since Channing's day. We have employed chloroform *à la reine*, intermittent nitrous oxide through long hours of vigil, twilight sleep, and various regional anesthetic procedures. About 20 years ago, Dr. Frederick Irving, our emeritus chief, explored the possibilities of the barbiturates, particularly for the stage of painful uterine contractions associated with dilatation and taking up of the cervix. This period may last for hours. Following this experience with barbiturates and scopolamine, paraldehyde, ether-in-oil, and other agents were explored to meet some of the undesirable reactions of the former.

The present Chief of the Boston Lying-in Hospital and Professor of Obstetrics at the Harvard Medical School, is Dr. Duncan E. Reid. He is carrying on this same noble tradition of taking all steps consistent with safe practice, so that every woman may bear her baby with a minimum of pain. To strive for safer practices of obstetrical amnesia, analgesia, and anesthesia is a challenge and responsibility all serious-minded physicians must recognize.

We have been impressed with the fact that our most successful methods have included the use of scopolamine. Its advantages and disadvantages may help point the direction for future exploration of newer synthetic agents that would find a wider application in obstetrical practice.

We believe scopolamine possesses the following advantages:

- (1) It is the best amnesic agent available.
- (2) Clinically, it produces psychic sedation with its accompanying relief of apprehension and anxiety, along with dreamless sleep.
- (3) It produces dryness of mucous membranes of the respiratory tract, often to a degree better than that produced by atropine and the newer available synthetics. A dry, unobstructed respiratory tract needs no emphasis here.
- (4) It can relieve bronchospasm and laryngospasm under certain circumstances.

- (5) It has the capacity of preventing some of the untoward physiological reactions, *e.g.*, the vagal type of carotid sinus syndrome (bradycardia, lowered arterial tension, and decreased pulse pressure).
- (6) It may increase maternal respiration.
- (7) It is more effective than atropine in combating respiratory depression produced by the more potent centrally acting narcotics.
- (8) Clinically, we have observed no ill-effects on fetal or neonatal functions.
- (9) It possesses a reasonably wide margin of safety.
- (10) Scopolamine can be administered by any route. The duration of its desirable clinical effects in obstetrics is about two hours.

We have found its chief disadvantages in obstetrics to be:

- (1) Its lack of sufficient analgesic potency. It has an effective psychological component in removing the fear-element, yet it lacks sufficient subcortical "blocking" effect at the lower integrating levels for pain.
- (2) It produces undesirable side reactions of excitement and muscular overactivity.
- (3) It may produce some other less frequent undesirable side effects, *e.g.*, edema of eyelids, lips, or uvula.

Can a drug be synthesized that possesses the advantages of scopolamine and yet lacks its disadvantages? Until such time, other drugs, including some of the newer synthetic analgesics, deserve careful exploration in the field of clinical obstetrics.

Recently, we have been limiting the barbiturates to the early period of labor for their hypnotic effect. As labor progresses, we have employed quite successfully the combination of scopolamine and apomorphine. The apomorphine is administered in subemetic doses to counteract some of the undesirable effects of scopolamine. It possesses some degree of potentiating the analgesic factor of scopolamine and helps considerably to maintain a quiet, efficiently functioning labor room. Thus far, in our hands, this approach has been the best solution to the problem from both the maternal and fetal standpoints of safety. In our apomorphine series,³ only 6 per cent of our newborns required resuscitative measures of some sort, and of this group but 1 per cent required early active resuscitative measures. All responded readily and maintained a desirable state of reflex irritability. In this same series, over 90 per cent of the mothers had satisfactory amnesia-analgesia and the labor room was quiet and peaceful on an active service.

We are now in search of a drug which will retain the merits of subemetic doses of apomorphine but increase the analgesic effect of scopolamine without diminishing the safety for mother and baby. At present, we have selected Demerol for study. Of all the newer synthetic analgesics more is known, probably, about the actions of Demerol and, further

more, we have had some little experience with this drug. In a previous study⁴ reported from the Boston Lying-in Hospital, it was noted that successful amnesia-analgesia occurred in about 70 per cent of the mothers. In this same series with Demerol, about 16 per cent of the newborns required active resuscitative measures at birth. Of this group, more than 3 per cent required immediate active resuscitative measures. But what is more important is the fact that an even larger group of newborns failed to maintain satisfactory reflex irritability even though adequate resuscitative measures were instituted early.

Brief Report on Current Use of Demerol

At present, our efforts are directed at exploring the possible merits of much smaller dosages of Demerol than heretofore reported. The present study consists of approximately 100 unselected primiparous patients at the same hospital. A test group were tried on 10-milligram doses with each two-hour administration of scopolamine, adding further 10-milligram intermediate doses of Demerol as indicated. This procedure resulted in improved reflex irritability of the newborn but with a much higher incidence of restlessness and excitement of mothers, poorer amnesia-analgesia, and a less quiet labor room.

More recently, we have been employing Demerol in 20-milligram doses with each administration of scopolamine, along with intermediate doses of Demerol in 20-milligram units as indicated. I must repeat that this is a report of work in progress, and the first consideration is safety for mother and baby. The facts in this study are as follows:

- (1) 47 per cent of the mothers had absolute amnesia. 43 per cent vaguely remembered some isolated incident. Consequently, satisfactory amnesia occurred in 90 per cent of the mothers. In this current series, 10 per cent of mothers failed to have amnesia. It must be emphasized that, in all our studies, the main objective is to produce satisfactory amnesia so that the patient has no memory of events from the time of administration of the drugs until she is awake in bed in the ward, *i.e.*, so that there is no memory whatever of labor. Of course, this is an affair for the hospital and not for home deliveries.

The medication is usually started in these primiparous patients when the cervix reaches 2 fingers dilatation and uterine contractions are recurring every 3-5 minutes and are of 40-45 seconds' duration.

- (2) Excitement occurred in 45 per cent of this series of mothers. The labor room is definitely not as quiet as with the apomorphine-scopolamine series. What is even more important, the individual mother is less apt to mobilize her expulsive efforts as effectively as the similar amnesic-analgesic mother does under the effects of scopolamine-apomorphine.

- (3) Emesis occurred in 13 per cent of this series. No other undesirable side effects were noted. There was no characteristic influence on maternal vital functions, *i.e.*, on respiration, pulse rate, or blood pressure. There was no characteristic effect on uterine rhythm, rate, or force of contraction.
- (4) The average length of labor was 11.4 hours. This is about the same duration as with the scopolamine-apomorphine series. Incidentally, the extremes in the present series varied from 1 to 43 hours. In our experience, the average unmedicated primiparous patient requires about 16 hours. The mothers in this series varied in age from 16 years to 34 years. Most of the cases were in the early half of their third decade of life. Their gestational age varied from 36 to 48½ weeks; the largest group were of 40 ± 2 weeks. Their weights varied from 116 to 211 lbs. and their heights from 51 to 68+ inches. Other factors influencing maternal reflex irritability varied correspondingly.
- (5) The blood loss was normal, *i.e.*, under 200 cc. in 87 per cent of the cases. 8.6 per cent of the cases had moderate hemorrhage and 4.4 per cent of this series had severe blood loss, *i.e.*, 400 cc. or over. Most incidents were explained on the basis of bleeding from the episiotomy or other operative wound, or lacerations or trauma during actual delivery.
- (6) 95.7 per cent of the cases were delivered vaginally, while 4.3 per cent were delivered by abdominal hysterotomy, after a test of labor of approximately 10 hours.
- (7) There were no maternal or neonatal deaths in this series. The average period of hospitalization was 10 days. In this series, there was one stillbirth. This occurred in a 22-year-old mother of 152 lbs. weight and 62 inches height, of 44 weeks' gestational age, having 9½ hours of labor, with uneventful prenatal course. She was delivered by low forceps extraction employing a medio-lateral episiotomy. The cord was found once around the neck, not too tightly. The immediate cause of death was undetermined. The pathologist's report was "congestion of lungs with petechial hemorrhages; congestion of viscera; dilatation of right ventricle; intrauterine asphyxia." The mother had no memory of labor but for vague recollection of actual delivery; she was delivered under low spinal analgesia without untoward effects of this anesthetic technique. What role Demerol may have played in this case will remain a mystery.
- (8) Immediate spontaneous respiration occurred in 29 per cent of the infants in this series; 46 per cent of the infants had slight delay of spontaneous respiration. Thus, 75 per cent respired satisfactorily at birth. 22 per cent of this series were resuscitated with comparative ease, and 1.5 per cent with great difficulty. Thus, in this series 23.5 per cent required resuscitative measures at de-

livery. Once resuscitated, this group of infants were not as markedly depressed as the original series reported under Demerol from this institution. With but few exceptions, all mothers and babies were discharged well on the 10th *post-partum* day. In a control series of unmedicated patients reported from this institution, 1.9 per cent of living full-term infants did not breathe immediately as soon as they were born. With the scopolamine-apomorphine series, 6 per cent of the infants required resuscitative measures at birth, but in this Demerol series 17.5 per cent more infants required resuscitation at birth.

- (9) The average case received four doses of Demerol, *i.e.*, 80 milligrams. The doses per case varied from 1-11, *i.e.*, 20 to 240 milligrams. The average interval between the first dose of Demerol administered and the time of delivery varied from 5-7 hours. The return to consciousness following delivery, on the average, was from 3-4 hours.
- (10) The babies were independently examined by the pediatric staff on the day of delivery, as well as on the day of discharge.

Conclusions

Certain impressions have been gained from the current studies of this synthetic analgesic agent:

- (1) Demerol is not an amnesic agent. This impression is gained, also, from a study of patients in whom Demerol alone was the sole pre-anesthetic agent.
- (2) Thus far, we have not observed excitation when Demerol alone was employed. Furthermore, Demerol will counteract the excitation of scopolamine when used in sufficient dosage.
- (3) Demerol will provide a limited amount of psychic sedation for the obstetrical patient in the dosage range currently employed. It will produce less depressing effects than, *e.g.*, morphine or its derivatives, in comparable dosages on the vital functions.
- (4) If a greater degree of analgesia and quietude of labor room is sought for, as compared to apomorphine, then the results are attained at a greater degree of depression of reflex irritability of the newborn.
- (5) Because of its ability to counteract some of the undesirable effects of scopolamine, its greater analgesic potency, its relative freedom from serious toxic reactions, and its availability for parenteral use, we shall continue to explore its possibilities to meet our problems in obstetrical amnesia, analgesia, and anesthesia for the mother if no further hazards are added to the baby.
- (6) There is room for further development of analgesic drugs for obstetrical purposes providing they are safer for mother and baby.
- (7) There is ample room for further development of scopolamine-like drugs, which may be of even greater value in meeting our require-

ments in obstetrics for more effective amnesia and analgesia for the mother without added risk for the baby.

Finally, the enthusiasm and prevalent interest in these problems offer even greater hope for the future development of still better and safer agents and methods of relieving the pains of childbirth.

References

1. HERSHENSON, BERT B. 1948. Premedication and anesthesia in obstetrics; practical aspects. *Anesthesiology* 9: 73-85.
2. CHANNING, WALTER. 1848. *A Treatise on Etherization in Childbirth*. Wm. D. Ticknor & Company. Boston.
3. HERSHENSON, BERT B., & E. R. BRUBAKER. 1947. Scopolamine and apomorphine in labor. *Am. J. Obst. & Gyn.* 53: 980-995.
4. SCHUMANN, W. R. 1944. Demerol (S-140) and scopolamine in labor. *Am. J. Obst. & Gyn.* 47: 93-104.

PRELIMINARY EXPERIENCES IN THE USE OF SOME OF THE NEWER ANALGESICS IN THE RELIEF OF PAIN DUE TO CANCER

By R. W. HOUDE, L. H. RASMUSSEN, AND J. S. LADUE

*Memorial Hospital for the Treatment of Cancer and Allied Diseases
and Sloan-Kettering Institute for Cancer Research, New York, N. Y.*

Introduction

THE purpose of this paper is to report the results of a clinical trial of four synthetic analgesics in patients who had severe pain consequent to inoperable cancer. The physician caring for the patient with inoperable cancer is almost always confronted by the need for measures which will control pain without keeping the patient at or near narcosis. Although roentgen and even chemotherapeutic treatment may control pain for variable periods of time, the disease all too frequently progresses beyond or becomes refractory to such measures. Specific surgical procedures such as chordotomy are often impractical and not always successful in their aim. The judicious use of analgesics then becomes, together with superficial psychotherapy, the only available means of combating mentally and physically crippling pain. Known analgesics such as the salicylates, meperidine, morphine and its analogues, often leave much to be desired because of unfortunate side effects, increasing tolerance, and drug addiction.

The ideal analgesic should not only control pain without sedative action, but perhaps should also produce a mild euphoria. The search for such a drug has been, and is still being, diligently conducted by chemists and pharmacologists and has resulted in the synthesis of several promising drugs.

Four chemically different analgesics were supplied to us for study of the problem of pain in patients with inoperable cancer admitted to Memorial Hospital. Two piperidine derivatives, NU-896 and NU-1196; methadone (Adanon hydrochloride or 10820), a drug made available in Germany at the end of World War II¹; and Metopon (methyl-dihydro-morphinone)² were investigated. The chemical formulae are given in FIGURES 1 to 3 (note the similarity between meperidine, NU-896, NU-1196, and methadone, as well as the close relationship between Metopon and morphine).

Metopon and methadone have been shown to have analgesic properties comparable to those of morphine.^{2, 6, 7, 9, 10} No clinical studies of the two piperidine compounds have, as yet, been reported. However, animal experiments show that the analgesic indices of NU-896 and NU-1196 approximate that of methadone; the relative safety of NU-896 is 4.8 times that of meperidine, while NU-1196 is 1.6 times safer than meperidine.^{4, 11}

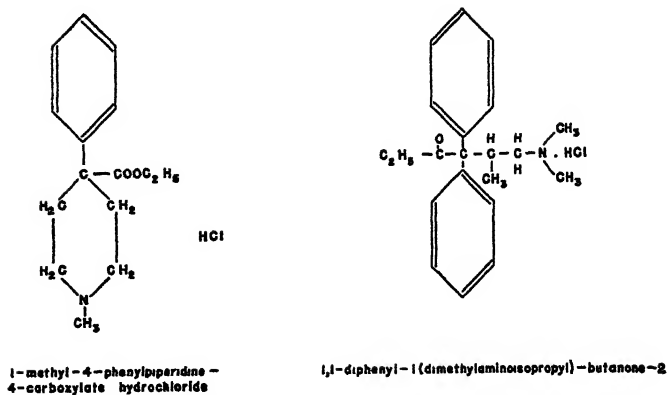


FIGURE 1. Formulae for meperidine hydrochloride (left) and methadone hydrochloride (right).

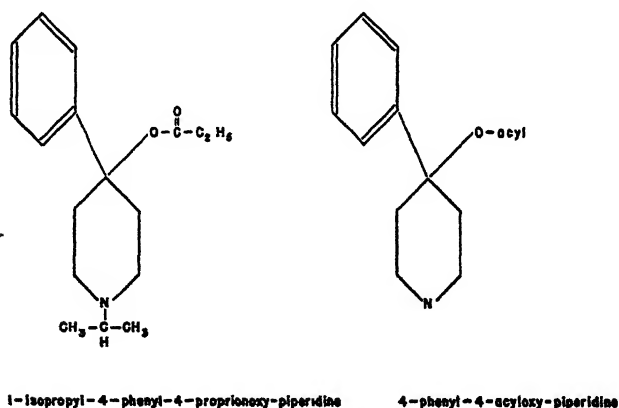


FIGURE 2. Formulae for NU-896 (left) and NU-1196 (right).

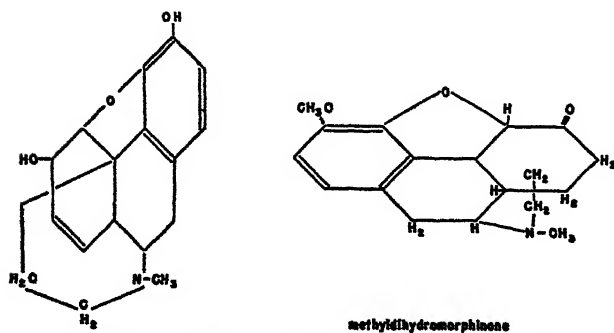


FIGURE 3. Formulae for morphine (left) and Metopon (right).

(TABLE 1.) These results suggested that these drugs were worthy of clinical trial.

TABLE 1
ACTIVITY AND SAFETY MARGIN OF ANALGESICS GIVEN SUBCUTANEOUSLY
(ANIMAL STUDIES)

<i>Drug</i>	<i>LD₅₀</i> <i>mg./kg.</i>	<i>AD</i> <i>1.40 mg./kg.</i>	<i>Safety</i> <i>margin</i>	<i>Relative safety</i> <i>(meperidine=1)</i>
NU-896	65	3.0	21.6	4.8
NU-1196	50	6.3	8	1.6
Methadone	45	3.4	13	2.6
Meperidine	250	50	5	1.0

Method of Study

The patient with severe pain due to cancer might, on first thought, appear to be an ideal subject for the study of the analgesic properties of any drug, but brief consideration will suffice to show that this is not the case. Most of these patients are apprehensive and depressed, and their reaction to pain is unpredictable. Many of them have several modalities of pain varying in intensity from time to time depending upon minor changes in position, the taking of food, or the elimination of excreta. Spontaneous exacerbations and relief of pain occur at unpredictable intervals. Our studies confirm the reported observation that data obtained on the pain response of animals or normal human subjects cannot be duplicated in patients who have experienced repeated painful stimuli over long periods of time.³ Moreover, it is not feasible to continue to administer relatively ineffective drugs to patients with severe, unrelenting pain and hence the extent of such studies must be limited. Despite these uncontrollable factors, we felt that a clinical investigation of these newer analgesics might demonstrate the usefulness of one or all of them in different situations.

TABLE 2 shows the disease processes responsible for pain in the 60 patients studied and the drugs employed in each situation. In TABLE 3, the number of patients presenting different types of pain problems and the drugs used in each instance are listed. Many patients had more than one type of pain but, for simplification, an arbitrary classification based upon the localization of pain, as well as the nature and extent of the neoplastic process, was adopted. Thus, these 60 patients presented 117 different pain problems depending upon the extent, extension, and method of production of the pain.

Each patient was evaluated as to the type and extent of his disease, his nutritional and psychosomatic status, as well as his experience of and pattern of reaction to pain (TABLE 4). Daily examinations for systemic change were made. Observations as to the onset, height, and duration of the patient's reaction to pain were made at frequent intervals both after giving the trial analgesic and after administration of the

TABLE 2

CLASSIFICATION OF PATIENTS AS TO DIAGNOSIS AND ANALGESIC GIVEN

<i>Diagnosis</i>	<i>Metopon</i>	<i>NU-896</i>	<i>NU-1196</i>	<i>Methadone</i>
Lymphoblastomas	1	3	3	11
Melanoma	—	1	1	—
Sarcomas	—	—	—	2
Carcinoma:				
Head and neck	3	—	1	13
Breast	1	—	1	5
Lung	—	1	—	2
Gastrointestinal	—	1	2	3
Rectal	—	—	1	2
G-U	—	—	—	3
Gynecological	—	—	1	2
Abdominal, primary?	—	—	—	2
Bone tumors:				
Multiple myeloma	—	—	—	1
Osteogenic sarcoma	—	—	—	1
Endothelioma	—	—	—	1
Metastatic, primary?	—	—	—	2
<i>Total:</i>	5	6	10	50

TABLE 3

CLASSIFICATION OF PATIENTS AS TO PAIN AND ANALGESIC GIVEN

<i>Localization of pain</i>	<i>Metopon</i>	<i>NU-896</i>	<i>NU-1196</i>	<i>Methadone</i>
Intracranial	—	1	1	4
Cord, root or nerve trunk	1	3	3	14
Visceral	2	4	8	20
Bone	3	1	3	25
Integumental:				
Pruritus	—	2	1	8
Ulceration	1	—	2	10
	7	11	18	81

standard analgesics employed for comparison studies. At first, an effort was made to obtain more objective data by measuring pain thresholds with the Wolff-Hardy-Goodell apparatus according to accepted techniques.³ It was exceedingly difficult to train these patients, since most of them were desperately ill. Few could cooperate when their pain was severe and, in many, basal conditions were unobtainable because of fever, cardiac, respiratory, or gastrointestinal distress. In the few that we were able to train, the results were so frequently inconsistent that it was impractical to continue the procedure. For example, one patient with extensive metastatic melanoma with cord lesions and radicular pain achieved an elevated pain threshold (250 to 360) after one of the analgesics but complained of the return of pain in the region of his disease at the height of the pain threshold elevation.

All of the patients studied had had pain for weeks or months and all had received several analgesics. In general, the salicylates, codeine and aspirin, meperidine, Pantopon, and morphine were used in that order

TABLE 4
METHOD OF STUDY

Preliminary data:

Age	Diagnosis	Nutritional status	Psychosomatic status
-----	-----------	--------------------	----------------------

Data obtained before and at intervals after giving analgesic:

Pain threshold as measured by Wolff-Hardy-Goodell apparatus

Type, location and radiation of pain

Pattern of reaction to pain (apprehensive, fear, depression, etc.)

Subjective change:

Onset of effect

Height of effect

Duration of effect

Criteria used in evaluating pharmacological action:

NEURO-MUSCULAR SYSTEM

Vegetative:

Temperature (rectal)

Skin blushing or blanching

Dermatographia

Cerebral:

Alertness

Apprehension

Depression

Euphoria

Drowsiness

Sleep

Dizziness

Headache

Sensory:

Paresthesia

Numbness

Hyperesthesia

Motor:

Effect on activity

Coordination

Special senses:

Pupil size

Reaction

Diplopia

Nystagmus

Tinnitus

Blurring

Acuity of hearing

OTHER SYSTEMS

Cardiovascular:

Heart rate and rhythm

Blood pressure

Respiratory:

Rate and depth

Effect on cough reflex

Gastrointestinal:

Nausea or vomiting

Abdominal distension

Cramps

Diarrhea

Constipation

Genito-urinary:

Pain

Burning

Atony

Frequency

for ascending severity of pain. Control observations were made in these patients for periods of from two to seven days before administration of any of the test drugs. When their pain was more or less controlled by one or more of the above listed analgesics, the trial drugs were then given first in the minimal recommended therapeutic doses and then in increased dosages as indicated. The dosage of NU-896 was 7.5 to 15 mg. by hypodermic injection; that of NU-1196, 15 to 40 mg. orally and hypodermically; of methadone, 2.5 to 20 mg. orally and hypodermically; Metopon was given by mouth only in doses of 3 to 18 mg. At first, all these drugs were given at four- to six-hour intervals and the dosages increased as the need for greater relief was manifested. If any degree of alleviation of pain was achieved, however, the interval was shortened as indicated by the duration of analgesia. When effective, the trial drugs were used over as long a period of time as possible in order to determine tolerance and toxicity. When the trial drug proved ineffective, the control analgesic was re-administered and re-evaluated. The various test drugs were not compared with one another, except in a few instances, as

they were obtained at different times during the period of investigation, and because many of the patients flatly refused to continue with a drug which they, or their attending physician, knew to be ineffective. We were unable to judge the addicting properties or withdrawal symptoms of these drugs since all the patients required and were given other analgesics when the trial preparation was withdrawn.⁸

Results

TABLE 5 shows the results obtained with these four drugs when compared according to the various types of pain. The degree of analgesia was arbitrarily classified as 1, 2, 3, or 4. Grade 1 indicated that the patients derived negligible analgesia. Patients with Grade 2 relief stated that the pain was somewhat less severe but that they were still definitely uncomfortable. Patients with Grade 3 relief were comfortable and able to carry on their activities, although when asked they stated that they were still aware of some dull pain. Grade 4 indicated that the patients were not conscious of any pain.

Of the five patients receiving Metopon, only one, a patient with a large radiation ulcer of the cheek, obtained significant relief of pain. This drug was uniformly unsatisfactory in doses up to 18 mg. given every two to three hours by mouth in the other patients with nerve trunk, visceral, and bone pain, and either the patients or their physicians demanded that they be placed on other medication. However, all of these patients had received other analgesics, and all patients but the one who did derive satisfactory analgesia had received morphine or Pantopon. Since Metopon was not available for clinical investigation and was difficult to obtain, our series is necessarily small and therefore not conclusive.

None of the patients receiving NU-896 obtained satisfactory analgesia. Two patients experienced Grade 2 relief of pain, and one with severe pruritus (Hodgkin's disease) some relief of itching. In none of these cases, however, was this drug as satisfactory as the previous medication.

Grade 3 to 4 analgesia was obtained in only two of ten patients receiving NU-1196. In a third patient, nerve trunk pain was relieved but visceral pain which developed subsequently was not. Thus, in only three patients with four pain modalities was satisfactory analgesia achieved.

Thirty-three of fifty patients who received methadone obtained Grade 3 to 4 analgesia. When grouped according to types of pain, as in TABLE 5, 59 per cent of the pain problems were satisfactorily controlled by this preparation (3 to 4 relief). When patients with pruritus were excluded, this percentage was raised to 63 per cent.

In TABLE 6, the time of onset and average duration of action of the drugs is noted and compared as to the route of administration. The onset of action was, in general, 15 to 30 minutes sooner when the drug was

TABLE 5
DEGREE OF ANALGESIA OBTAINED

Localization of pain	Metopon				NU-896				NU-1196				Methadone			
	(1)	(2)	(3)	(4)*	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
Intracranial	-	-	-	-	1	-	-	-	1	-	-	-	1	-	3	-
Cord, root, nerve trunk	1	-	-	-	2	1	-	-	1	1	1	-	1	4	9	-
Visceral	2	-	-	-	3	1	-	-	5	1	2	-	6	4	8	2
Bone	3	-	-	-	1	-	-	-	2	1	-	-	3	7	14	1
Ulceration	-	-	1	-	-	-	-	-	-	1	1	-	0	1	8	1
(Pruritus)	-	-	-	-	1	1	-	-	1	-	-	-	4	2	1	1
Total	6	0	1	0	8	3	0	0	10	4	4	0	15	18	43	5
Total, per cent		0	14	0		27	0	0		22	22	0		22	53	6
Pain relief, per cent				14				27				44				81
Grade (3) and (4) relief, per cent				14				0				22				59

* Pain relief: (1) negligible, (2) some, (3) good, (4) complete.

TABLE 6

COMPARISON OF EFFECTIVE ANALGESIA WITH DRUGS TESTED

		<i>Dose (mg.)</i>	<i>Onset (minutes)</i>	<i>Average duration (range) (hours)</i>
Metopon	(oral)	3-18	30	2*
NU-896	(hypo)	7.5-15	20-35	1½ (1-2)
NU-1196	(oral)	20-40	30-60	1½ (1-3)
	(hypo)	20-40	15-30	2½ (½-5)
Methadone	(oral)	2.5-10	30-60	4 (1-12)
	(hypo)	5-20	15-30	4 (1-8)

* One case.

given hypodermically than when given by mouth. The average duration of analgesia in individual patients is indicated in parentheses. The overall average for all the patients is indicated by the preceding figure in this column. The single patient in this series who responded to Metopon noted appreciable analgesia for two hours only. Those who noted any effect from NU-896 noted pain relief for an average of one to two hours, or an overall average of one and a half hours. NU-1196 given orally showed an average duration of analgesia of one and a half hours, with a range of from one to three hours. When given hypodermically, the average duration of analgesia was two and a half hours, with a range of from one to five hours. Methadone, either by mouth or hypodermically, produced analgesia of about four hours' duration. The average duration in individual patients ranged from one to twelve hours by mouth and one to eight hours by hypodermic injection. It is to be noted, however, that those who received the drug parenterally were as a rule more ill and apprehensive and, as far as could be judged, were experiencing more severe pain.

When the effects of oral and hypodermic administration of NU-1196 and methadone were compared, as in TABLE 7, both routes appeared equally effective. However, this, too, was more apparent than real and resulted from the selection of patients. Patients who had been receiving analgesics parenterally prior to the giving of the trial analgesics were usually given the new drug by the same route so that they would not be aware of any change. This was done in order to minimize the psychological effect of a "new drug" on the patients' experience of pain. In the few comparisons that were made, patients receiving hypodermically administered medication did not in any instance derive satisfactory analgesia with orally administered analgesics even though the dose was often twice as much. However, in two of three patients who were given NU-1196 orally with no appreciable effect, subsequent hypodermic injections in the same dosage were effective. When methadone was given subcutaneously, approximately one-half the dosage was required to produce the same effect as when it was given orally.

In TABLE 8, a comparison of the relative analgesic efficacies of the trial and standard drugs is presented. A comparison of the effect of the

TABLE 7
COMPARISON OF ANALGESIA BY ORAL AND HYPODERMIC DOSES

Localization of pain	NU-1196				Methadone							
	Oral				Hypo				Oral			
	(1)	(2)	(3)	(4)*	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
Intracranial	-	-	-	-	1	-	-	-	1	-	3	-
Cord, root, nerve trunk	-	-	1	-	1	1	5	-	1	3	4	-
Visceral	2	1	-	-	5	-	2	-	3	1	6	1
Bone	-	-	-	-	2	1	-	-	1	2	11	1
Ulceration	-	-	-	-	-	1	1	-	-	1	4	-
(Pruritus)	1	-	-	-	-	-	-	-	1	1	1	-
Total	3	1	1	-	9	3	3	-	6	6	16	2
Total, per cent	-	20	20	-	-	20	20	-	-	19	52	10
Pain relief, per cent	-	-	-	40	-	-	-	-	-	21	54	4
Grade (3) and (4), per cent	-	-	-	20	-	-	-	-	-	-	-	79
	-	-	-	-	-	-	-	-	-	-	-	58

* Pain relief: (1) negligible, (2) some, (3) good, (4) complete.

TABLE 8

COMPARISON OF ANALGESICS

(Percentage preference for drug listed in column at left)

<i>Trial analgesic</i>	<i>Dose (mg.)</i>	<i>Morphine</i>	<i>Pantopon</i>	<i>Demerol</i>	<i>Codeine and aspirin</i>
		<i>4 pts.</i>	<i>3 pts.</i>	<i>1 pt.</i>	<i>3 pts.</i>
Metopon	3-9	0% (10-15)*	0% (10-20)	100% (50)	0% (65)
	12-18	0% (15)	0% (20)	—	—
NU-896	7.5	<i>2 pts.</i>	<i>2 pts.</i>	<i>2 pts.</i>	—
	15	—	—	—	—
		0% (10)	0% (10-20)	50% (100)	—
NU-1196	20	<i>7 pts.</i>	<i>6 pts.</i>	<i>5 pts.</i>	<i>8 pts.</i>
	30-40	0% (15)	0% (10)	50% (50-100)	50% (65)
		0% (15)	0% (20-30)	100% (50-100)	50% (65)
Methadon	2.5-7.5	<i>20 pts.</i>	<i>14 pts.</i>	<i>30 pts.</i>	<i>32 pts.</i>
	10-20	15% (10-15)	0% (20)	50% (50-100)	80% (30-65)
		45% (10-100)	43% (10-20)	94% (100)	93% (30-65)

* Figures in parentheses represent doses in mg. of drugs used for comparison.

trial analgesic with one or more of the opiates in general use for pain relief was made in all the patients in this series. In doses of 3 to 19 mg. by mouth, Metopon was inferior to morphine, Pantopon, and codeine and aspirin. However, the one patient cited above preferred 3-mg. doses of Metopon to 50 mg. of meperidine, but not to oral doses of 5 mg. of methadone or 10 mg. of morphine. NU-896 was inferior in dosages of 7.5 to 15 mg. to 10 mg. doses of morphine or Pantopon, although 15-mg. doses of NU-896 were preferable to 100 mg. of meperidine in one of two patients. NU-1196 in doses of 20 to 40 mg. was less satisfactory than 15 mg. of morphine in seven patients, and than 10 to 30 mg. of morphine in six patients. Two of four patients preferred 20-mg. doses of NU-1196 to 50 to 100 mg. of meperidine, and one patient preferred 40 mg. of NU-1196 to 100 mg. of meperidine. In two of four patients, 20 mg. of NU-1196 was more satisfactory than 65 mg. of codeine together with 650 mg. of aspirin, but in two of four patients 30 to 40 mg. of the test drug was less satisfactory.

Methadone in 2.5 to 7.5 mg. doses was preferred to 10 to 15 mg. of morphine in only two of thirteen patients (15%), but in 10 to 20 mg. doses was preferred in eight of eighteen (45%). Two patients who were addicted to morphine and were receiving 292 and 455 mg. of morphine daily in doses of 92 to 100 mg. every 4-6 hours were not well controlled with methadone in 10-mg. doses, although, upon the insistence of the physicians attending these patients (who were approaching terminus), this was not pursued to a conclusive test. Six of fourteen patients (43%) preferred methadone in 10 to 20 mg. doses to 20 mg. of Pantopon, but in two of these, methadone in 5-mg. doses was unsatisfactory. Six of twelve patients were better controlled with 2.5 to 7.5 mg. of this drug than with 50 to 100 mg. of meperidine, and eighteen of nineteen (94%) with 10 to 20 mg. of methadone than with 100 mg. of meperidine. Twelve of fifteen patients preferred 2.5 to 7.5 mg. of methadone to 30 to 65 mg. of codeine (usually combined with aspirin), and in sixteen of seventeen (93%) 10 to 20 mg. doses were preferable to these same amounts of codeine and aspirin. The "preference" for a given analgesic was determined not only by the patient's stated opinion but also by the observers' estimation of the degree and duration of pain relief, and the presence or absence of side effects.

The majority of patients at Memorial Hospital are admitted for surgery. Our aim, however, was to study analgesics in those in whom the relief of pain was the primary consideration, and since the majority of patients in our series had inoperable cancer, their admissions to the hospital were scattered over the period of study. This, plus the fact that all the trial drugs were not made available for use at the same time, interfered with attempts to compare these drugs with each other. The comparisons that were made indicated that methadone was superior to the other three, but there were too few of these to be considered significant and, hence, these data are not presented in table form.

TABLE 9
INCIDENCE OF SIDE EFFECTS

	<i>Metopon</i>		<i>NU-896</i>		<i>NU-1196</i>		<i>Methadone</i>	
<i>Dosage (mg.):</i> —	<i>3-9</i>	<i>12-18</i>	<i>7.5</i>	<i>15</i>	<i>20</i>	<i>30-40</i>	<i>2.5-7.5</i>	<i>10-20</i>
Fall in blood pressure								1
Respiratory depression								1
Dizziness							1	1
Nausea or vomiting								2
Diaphoresis						1		
Paraesthesias								
Sedation*						3	2	2
Suppression cough*								2
<i>Total side effects</i>	0	0	0	0	0	4	3	9
<i>Side effects excluding *</i>	0	0	0	0	0	1	1	5
<i>Per cent side effects excluding *</i>	0	0	0	0	0	10	2	10

* Excluding sedation and cough suppression as undesirable side effects.

TABLE 9 indicates the incidence of side effects of these drugs. No side effects were noted when Metopon or NU-896 was given. One patient who received 40-mg. doses of NU-1196 developed diaphoresis. Three patients receiving 30 to 40 mg. doses developed sedation to the point of hypnosis. One ambulatory patient receiving 5 mg. doses of methadone complained of dizziness and in two bed-ridden patients sedation was noted after 5 to 7.5 mg. doses. With doses of 10 to 20 mg., one ambulatory patient developed dizziness and in two patients the drug had to be discontinued because of nausea and vomiting. In two patients, moderate sedation was noted and in two, definite suppression of the cough reflex. Only one patient exhibited a fall in blood pressure and depression of the respiratory rate. This patient, who was dying of widespread reticulum cell sarcoma, was receiving 20 mg. of methadone every two to three hours and as much as 120 mg. per day, developed marked sedation, suppression of the cough reflex, a fall in blood pressure from 120-130/80 to 90-100/60-70, and a fall in respiratory rate from 24 to 12 per minute. No other toxic effects such as miosis, euphoria, skin rash, constipation, or changes in blood counts or blood chemistries which could be attributed to the drug were noted, although one patient received a total of 1195 mg. of methadone over a period of 127 days. Thus, toxic effects were observed in no patients receiving Metopon or NU-896, in 10 per cent of those receiving NU-1196, and in 12 per cent of those receiving methadone. It is interesting to note that none of the patients developed constipation while taking the trial analgesics. Two patients who had developed annoying constipation while taking opiates noted a return of normal bowel habits when given methadone.

Increased tolerance to an analgesic dose of NU-896 was striking and developed within two to three days in all the patients to whom this drug was given. The one patient who derived any pain relief from Metopon showed no increased tolerance but was observed for only two weeks.

Three patients who received NU-1196 for from two weeks to three months were noted to have attained an increased tolerance, developing within the first week. However, another patient received from 40 to 60 mg. of NU-1196 by hypodermic injection daily for four months without showing an increased tolerance. Only two of the patients given methadone required increasing amounts of this drug for relief.⁸ Six patients who received methadone for from eighteen to 127 days failed to display increased tolerance. Since none of these patients was ever without some analgesic drug, it is impossible to state whether or not withdrawal symptoms might have developed.

Summary and Conclusions

- (1) Four synthetic analgesics were given to sixty patients with persistent severe pain due to cancer and the efficacy of each preparation was estimated.
- (2) NU-896 did not produce satisfactory analgesia in any of the six patients to whom it was given.
- (3) NU-1196 gave effective relief of pain in two of ten patients.
- (4) Metopon gave adequate relief to one of five patients.
- (5) Methadone produced satisfactory analgesia in thirty-three of fifty patients.
- (6) 10 milligrams of methadone appear almost as effective as 15 milligrams of morphine.
- (7) 10 milligrams of methadone were almost uniformly preferable to 100 milligrams of meperidine or to 65 milligrams of codeine.
- (8) No side effects were noted with NU-896 and Metopon.
- (9) Side effects were noted in 10 per cent of the patients receiving NU-1196 and in 12 per cent of those receiving methadone.
- (10) Increased tolerance occurred in all patients receiving NU-896, in three of four patients who received NU-1196 for more than two weeks, and in two of ten patients receiving methadone for more than two weeks. It was not noted in the single patient receiving Metopon for a period of two weeks.

Bibliography

1. KLEIDERER, E. C., J. B. RICE, V. CONQUEST, & J. H. WILLIAMS. 1945. Report 981, U. S. Department of Commerce, Office of the Publication Board, Washington, D. C.
2. EDDY, N. B. 1939. Studies of carbazole derivatives. I. Aminocarbazoles. *J. Pharm. & Exp. Therap.* 65: 294.
1939. Studies of carbazole derivatives. II. Amino alcohols and derivatives of tetrahydrocarbazole. *Ibid.* 65: 308.
1939. The search for more effective morphine-like alkaloids. *Am. J. Med. Sci.* 197: 464.
3. WOLFF, H. G., J. D. HARDY, & H. GOODELL. 1940. Studies on pain. Measurement of effect of morphine, codeine, and other opiates on pain threshold and analysis of their relation to pain experience. *J. Clin. Invest.* 19: 659.

4. BATTERMAN, R. C. 1943. Clinical effectiveness and safety of a new synthetic analgesic drug, Demerol. *Arch. Int. Med.* 71: 345.
5. Report of the Council on Pharmacy and Chemistry of the American Medical Association. 1946. *J. A. M. A.* 132: 146.
6. SCOTT, C. C., & K. K. CHEN. 1946. The action of 1,1-diphenyl-1-(dimethyl-amino-isopropyl)-butanone-2, a potent analgesic. *J. Pharm. & Exp. Therap.* 87: 63.
7. ISBELL, H., A. WIKLER, A. J. EISENMAN, & K. FRANK. 1947. Effect of single doses of 10820 on man. *Fed. Proc.* 6: 341.
8. ISBELL, H., A. WIKLER, N. B. EDDY, J. L. WILSON, & C. F. MORAN. 1947. Tolerance and addiction liability of 6-dimethylamino-4-4-diphenyl-heptanone-3 (methadon). *J. A. M. A.* 135: 888.
9. KIRCHHOF, A. C., & N. A. DAVID. 1948. A clinical trial of amidone (dolophine), a new synthetic analgesic. *Anaesth. & Analg.* 27: 92.
10. TROXIL, E. B. 1948. Clinical evaluation of the analgesic methadon. *J. A. M. A.* 136: 920.
11. Hoffmann-La Roche, Inc. Personal communication.

November 30, 1948

AUREOMYCIN - A NEW ANTIBIOTIC*

Consulting Editor: J. H. WILLIAMS

CONTENTS

	PAGE
Aureomycin, a Product of the Continuing Search for New Antibiotics. <i>By</i> B. M. DUGGAR	177
The Pharmacology of Duomycin. <i>By</i> B. K. HARNED, RAYMOND W. CUNNINGHAM, MARY C. CLARK, RACHEL COSGROVE, CAROLINE H. HINE, WM. J. McCAULEY, EDGAR STOKEY, R. E. VESSEY, N. N. YUDA, AND Y. SUBBAROW	182
Bacteriological Studies of Aureomycin. <i>By</i> CLIFFORD W. PRICE, WM. A. RANDALL, AND HENRY WELCH	211
The Determination of Aureomycin in Serum and Other Body Fluids. <i>By</i> A. C. DORNBUCH AND E. J. PELCAK	218
<i>In Vitro</i> Studies with Aureomycin. <i>By</i> CAROLINE A. CHANDLER AND ELEANOR A. BLISS	221
Laboratory Studies with Aureomycin. <i>By</i> THOMAS FITE PAINE, JR., HARVEY SHIELDS COLLINS, AND MAXWELL FINLAND	228
Clinical Studies with Aureomycin. <i>By</i> HARVEY SHIELDS COLLINS, THOMAS FITE PAINE, JR., AND MAXWELL FINLAND	231
Studies on Serum Concentrations in Humans and Preliminary Observations on the Treatment of Human Infections with Aureomycin. <i>By</i> HARRY F. DOWLING, MARK H. LEPPER, LEWIS K. SWEET, AND ROBERT L. BRICKHOUSE	241
Use of Aureomycin on Some Experimental Infections in Animals. <i>By</i> PAUL A. LITTLE	246
Treatment of Experimental Infections with Aureomycin. <i>By</i> MORTON S. BRYER, EMANUEL B. SCHOENBACH, ELEANOR A. BLISS, AND CAROLINE A. CHANDLER	254
The Pharmacology and Clinical Trial of Aureomycin; A Preliminary Report. <i>By</i> EMANUEL B. SCHOENBACH, MORTON S. BRYER, AND PERRIN H. LONG	267
Aureomycin in Ocular Infections. <i>By</i> ALSON E. BRALEY AND MURRAY SANDERS	280
Action of Aureomycin against Experimental Rickettsial and Viral Infections. <i>By</i> SAM C. WONG AND HERALD R. COX	290
Aureomycin—A New Antibiotic with Antirickettsial Properties: Its Effect on Experimental Spotted Fever and Epidemic Typhus. <i>By</i> LUDWIK ANIGSTEIN, DOROTHY M. WHITNEY, AND JOE BENINSON	306
The Treatment of Lymphogranuloma Venereum and Granuloma Inguinale in Humans with Aureomycin. <i>By</i> L. T. WRIGHT, M. SANDERS, M. A. LOGAN, A. PRIGOT, AND L. M. HILL	318
Treatment of Q Fever in Man with Aureomycin. <i>By</i> E. H. LENNETTE, G. MEIKLEJOHN, AND H. M. THELEN	331

*This series of papers is the result of a Conference on Aureomycin held by the Division of Mycology of the Section of Biology, The New York Academy of Sciences, on July 21, 1948.
Publication made possible by a grant from Lederle Laboratories Division, American Cyanamid Company.

Copyright, 1948, by The New York Academy of Sciences

AUREOMYCIN: A PRODUCT OF THE CONTINUING SEARCH FOR NEW ANTIBIOTICS

By B. M. DUGGAR

Lederle Laboratories Division, American Cyanamid Company,
Pearl River, N. Y.

THE articles herewith presented discuss a new antibiotic from a new species of the Actinomycetes—these Actinomycetes constituting a group of fungi which might well be called ultra-molds. We are concerned with a species of the established genus *Streptomyces*, for which the name *Streptomyces aureofaciens* is being proposed. The significance of the specific name will soon be apparent. This is a soil organism which seems to have escaped previous publicity and domestication, as well as having evaded a pigeonhole in the accessions of mycological science.

The antibiotic (aureomycin) from this species of mold will be evaluated, in preliminary measure at least, in relation to the role which, at this stage of the studies in progress, it may promise to play safely in the treatment of disease. Academically, the antibiotic itself and the organism producing it are of very considerable interest, whatever may be the score as affecting human health or animal production. But we measure our interest in it, at the present time, primarily by the qualities which it may display as a systemic and an effective chemotherapeutic agent.

It was most fortunate that so soon as reasonable quantities of aureomycin became available and pharmacologically acceptable, it was possible to obtain the cooperation of so many recognized leaders in the fields of experimental, clinical, and other aspects of medicine and medical work to give it immediate hospital trial, thus gaining the weight of their judgments as to its future possibilities. The authors, as well as the titles of their papers, suggest the wealth of skill and experience, the variety of diseases available, and the character of the facilities employed in the observations and investigations which followed.

My purpose is, first, to be introductory but brief; secondly, to say something of the characteristics of the source species and of its antibiosis *in vitro*, especially pointing out or emphasizing the range of disease-inducing or other organisms which, by growth-inhibition (for one criterion), have demonstrated susceptibility to aureomycin, and likewise to suggest certain groups of organisms which are more resistant.

Before proceeding with a very few of the details referred to, permit me to make two acknowledgments. The first is this: In referring to any experimental work with either the proposed *S. aureofaciens*, or with the product aureomycin, in this preliminary record, it should be clear without further emphasis that the direction of the work by the late Dr. SubbaRow and his sustained encouragement are gratefully acknowl-

edged; likewise cordially acknowledged are the suggestions of co-workers, particularly the assistance of my associates, Misses Doris Dansby and Dorothy Evans. The second acknowledgment is of a different order. It is that I shall present today no history of antagonism, antibiosis, and aversion of micro-organisms nor any citation of important events characterizing either the present antibiotic period or the earlier ante-antibiotic period. It may be added that much has been said on antibiotic history in the Academy's conference room, and the record is available in its *Annals*. Nevertheless, there are a few general facts and figures which seem to require rehearsal for better orientation.

From digests and summaries appearing in 1946-47 it appears that, up to that time, there were recognized roughly 30-35 antibiotics, excellent or indifferent, having their origin in the metabolic products of the usual molds or filamentous fungi, and there were about the same number of such agents attributable to the bacteria. In considerable part the two groups of substances are unlike. Then there are the Actinomycetes, which the rank and file of the mycologists treated with static contempt—and the bacteriologists in large part kicked around like the legendary "houn'-dog." These last, the Actinomycetes, furnished, up to the time mentioned, about a dozen antibiotics, and typically these present few products resembling those of the bacteria. The total of these antibiotics or substances listed as such was then about 80.

It is likely that the present total is now more nearly 100, the result of feverishly active surveys and researches during this post-war period. Perhaps half of the recent product "finds" have resulted from studies in the Basidiomycetes, commonly known as fleshy, bracket, or punk fungi. About as many additions come from the Actinomycetes and the bacteria. Of this respectable galaxy, the industrial output of therapeutic agents is largely limited to penicillin and streptomycin, fumigacin and streptothricin among mold products being as it were tagged for further study as systemic agents. The status of the bacterial products is still largely uncertain partly because some are new arrivals, although gramicidin, tyrothricin and bacitracin have, perhaps, an established but limited place.

Few persons, I think, would wish to commit themselves very far on the outlook for the future. The guess might be ventured that the apparent failures to bring forward new antibiotics arising from filamentous molds may be less indicative than the silence seems to warrant. It may be that investigators have become sensitized to high potencies, as witnessed in the stepped-up yield of penicillin. It should not be overlooked that initially low potencies (relatively) have also characterized other products than penicillin. It may be guessed also that fame-seeking Actinomycetes are not in waiting around our doorsteps, and the cost of going after them and the disappointments to be achieved are some things to contemplate. Nevertheless, the promise, I would say, is more than adequate to justify continuance of the search.

Returning now to a limited characterization of the proposed *Streptomyces aureofaciens*, a significant point is this: At a certain stage in the growth of the colony of this fungus, or of a smear (an amalgamation of colonies), there is "typically" the production of a golden yellow pigment in the moist or hygrophorus substrate mycelium, hence the proposed specific name, *aureofaciens*, signifying a golden appearance. Moreover, the antibiotic is also faintly golden yellow, an added reason why this antibiotic from *S. aureofaciens* is designated *Aureomycin*.

Color production is commonly well developed in most strains of this organism isolated from the soil and grown on meat extract-asparagine-dextrose agar or on potato-dextrose agar, and on certain other media as well, notably on potato cylinders. Account is not taken here of certain "washed-out" strains and mutants. Grown on agar, the substrate mycelium of young colonies is practically hyaline at first, commonly becoming yellow in 2-3 days. The aerial or "dry" mycelium is white, that is, without color at first. Likewise, the first-formed spores are white, but the entire heavily sporing surface of a slanted agar culture gradually changes (in 5-7 days at 28° C.) through brownish gray to a dark, drab gray. At the same time most of the substrate mycelial color disappears. The reverse color of slants at its best is golden tan, later tawny. With continued incubation, or if refrigerated, the mycelial color is promptly dulled. Microscopically, the hyphae have the appearance of true fungous filaments, and the aerial ones break up into chains of spores in a manner characteristic of certain fungi.*

There are several species of *Streptomyces* characterized by the production of yellow pigment in the mycelium or in the culture medium—the latter the so-called soluble pigments. With the cooperation of Mr. C. W. Hesseltine a careful comparison has been made with the original description of each of these species and with living culture material so far as available. The following are included under the "yellow" species of *Streptomyces* with which comparisons have been made: *S. flaveolus*, *S. californicus*, *S. cellulosa*, *S. parvus*, *S. flavochromogenes*, *S. antibioticus*, *S. aureus*, *S. flavus*, *S. citreus*, *S. fulvissimus*, *S. alboflavus*, and others which are omitted for various reasons. This study has involved, of course, detailed consideration of those reactions termed "cultural characteristics," as also of certain special biochemical properties, all of which will be presented in the form of a technical diagnosis in a paper now being held for publication, to follow the printing of this report.

The present report would be wholly incomplete without an adequate statement of the *in vitro* antibiosis of this species. Besides many tests of all strains, a typical strain, A-377, has been tested, to show the sensitivity or resistance of each of more than 50 species of the usual or unusual assay organisms. The list of organisms tested includes pathogenic bacteria (human or animal), non-pathogenic bacteria, plant pathogens, and

*FIGURES 1, 2, 3, and 18 display the development of *S. aureofaciens* as far as it seems practicable to exhibit this aspect of its life history at the present time.

various filamentous fungi and yeasts. The procedure used was not merely the usual form of spectral test, but rather placements (colonies) of the growing A-377 (or related strains*) were grown for approximately 75 hours on agar plates appropriately flooded with the assay organism somewhat as in the case of the cup technique. The assay, however, involves the exposure of the flooding organism to the antibiotic directly as secreted by the growing colonies of *S. aureofaciens*. Just as a demonstration of the procedure refer to FIGURE 10, displaying *S. griseus* colonies (or placements), tested against *Trichophyton mentagrophytes* as the flooding, or assay, organism. This is used because the zones are much smaller than those of our A-377 and the differential areas are defined more nearly in the center of the plate. The diameter in mm of the zone of inhibition, under standard conditions, furnishes an approximate criterion of activity or potency of the antibiotic-furnishing strain, or conversely an index of sensitivity of the assay species. Eventually the activity of the strain may be compared with a standard. As exhibited on the plate, activity is related to inherent potency and concentration of the antibiotic and its rate of diffusion, as well as to other factors not here considered.

Twenty-two Kodachrome slides were made displaying inhibition zone assays, a number of which show "off-plate" or "practically off-plate" effects in relation to the assay organism †. Essentially all isolates, more or less directly from soil dilution cultures gave off-plate often spectacular values with the following gram-positive bacteria: *Bacillus cereus*, *Bacillus subtilis*, *Staphylococcus aureus*, *Streptococcus pyogenes*, and *Mycobacterium tuberculosis* (607), and with the following gram-negative ones, *Escherichia coli*, *Aerobacter aerogenes*, *Salmonella pullorum*, *Proteus vulgaris*, *Klebsiella pneumoniae*, *Neisseria catarrhalis*, and *Eberthella typhosa*. High values

* The induced mutant A 377-899 was furnished by my colleague Dr F J Backus

† These are here represented by FIGURES 4-12 and 14-24

FIGURES 1-12 (see opposite page)

FIGURE 1 Development of colonies of *Streptomyces aureofaciens* (A 377), growing on meat extract dextrose asparagine agar

FIGURE 2 Agar slant culture of *S. aureofaciens* (A 2020) sporing heavily

FIGURE 3 Close-up of colony of *S. aureofaciens* (A 2020) displaying particularly the early development of white aerial hyphae

FIGURE 4 Large inhibition zones in assay of zone values of *S. aureofaciens* (isolate A 271) against *Staphylococcus aureus*

FIGURE 5 Overlapping and off-plate (designated OP values)* zones of *S. aureofaciens* (isolate A 377) against *Staph. aureus* as assay organism

FIGURE 6 Practically, OP inhibition zones of *Penicillium chrysogenum* (strain Q 176) against *Staph. aureus*

FIGURE 7 OP inhibition zones of *S. aureofaciens* (isolate A 232) against *Mycobacterium tuberculosis* (No 607)

FIGURE 8 Partly overlapping inhibition zones of *S. aureofaciens* (isolate A 1884) an organism of characteristically weak potency in this group of isolates against *Mycob. tuberculosis* (No 607)

FIGURE 9 OP inhibition zones of *S. aureofaciens* (A 377) assayed against the fungus *Trichophyton mentagrophytes*. The scattered clumps are flakes of mycelium of the assay organism introduced with the spores, no growth arising from the flakes in 48 hours

FIGURE 10 Inhibition zones produced by *S. griseus* (strain A 1) against *Trichophyton mentagrophytes*

FIGURE 11 OP inhibition zones of *S. aureofaciens* (induced mutant A 377-899) against *Salmonella pullorum*

FIGURE 12 OP inhibition zones of *S. aureofaciens* (induced mutant A 377-899) against the organism *Phytophthora tumefaciens*, the causal bacterial agent of "crown gall" of plants



Streptomyces aureofaciens A277

Streptomyces aureofaciens
A2020



1

2

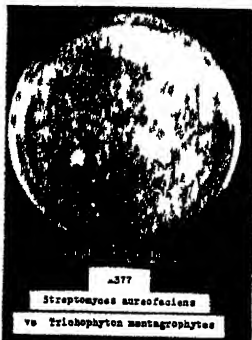
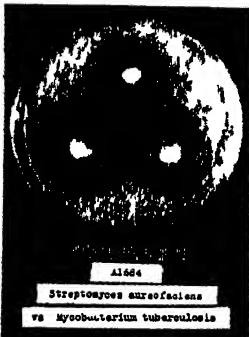
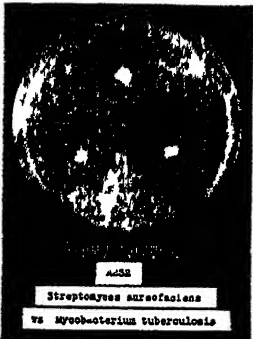
3



4

5

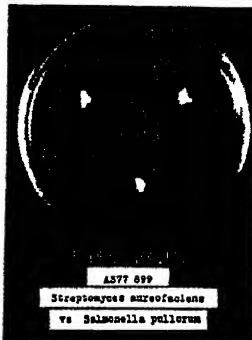
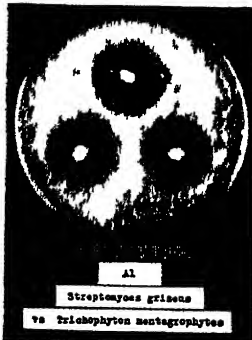
6



7

8

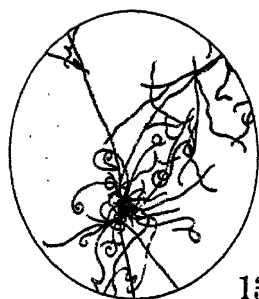
9



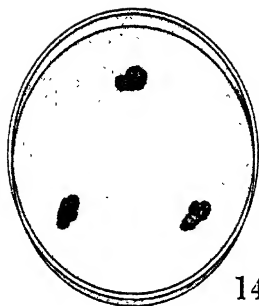
10

11

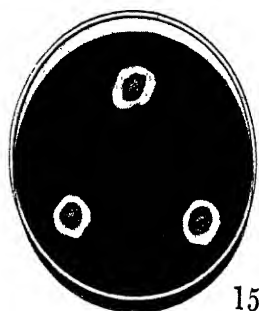
12



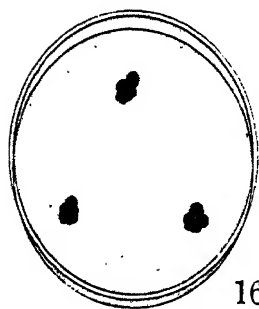
13



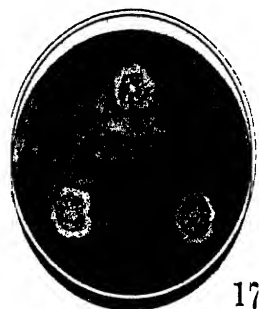
14



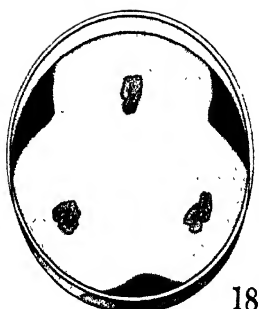
15



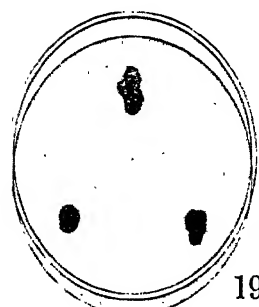
16



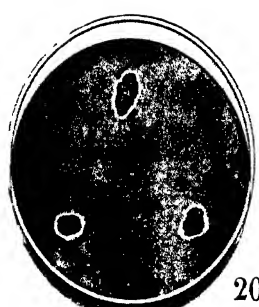
17



18



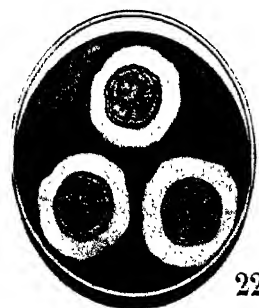
19



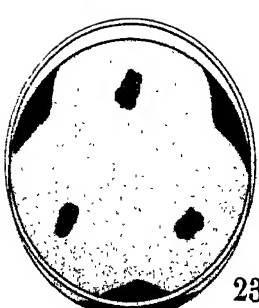
20



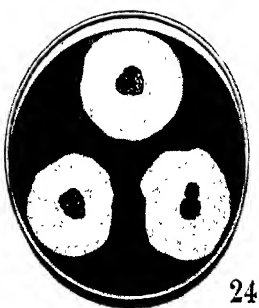
21



22



23



24

were also obtained against *Shigella gallinarum*, *Brucella abortus* and other species. Relatively resistant were *Pseudomonas aeruginosa* and *Serratia marcescens*, while the majority of the fungi and yeasts are strongly resistant. The figures display differences and relations not mentioned in the text. The photographic work on the original slides was done by Mr. L. H. McWilliam. The paintings, from which four of the slides were prepared, were the work of Mr. A. A. Jansson.

The culture of *Streptomyces aureofaciens* in shaken flasks and tanks has been satisfactory. The highly purified, crystalline product has been tested in broth at a favorable pH against a variety of organisms in order to determine more nearly absolute effects. These values confirm in considerable measure the expectations formulated as a result of the plate-zone tests, but they will not be presented here in view of the probability that other papers today will cover this aspect of the work. Tests made at a pH much above pH 7 have little point, since loss of the product by precipitation introduces one type of error, or, if the period of observation is shortened, the extent of growth, if any, is limited and the turbidimetric assay rendered uncertain.

Now, when all this demonstration of conspicuous activity has been displayed by plate zones and other types of assay, one is certain that there is a potent antibiotic present. It is in this sense only, at that stage, a desirable antibiotic. It must be demonstrated that the organism will give fair to high potencies when cultured in suitable liquid media, and the product obtained must assay accordingly when tested against some of the same microorganisms. The problems of isolation, purification, and crystallization of the antibiotic are at hand. Finally, there are the various types of *in vivo* activity, of stability, pharmacological suitability, and of efficiency for particular infections. These all will be treated in the following papers.

FIGURES 13-24 (see opposite page)

It is to be noted that there are three values in the intensity of the tones used in FIGURES 13-24, and the following explanation applies in all cases: (a) the dark tone (ink stipple) represents the fungus tested for antibiotic activity, that is, the "placement" organism present in each of the eleven figures indicated above; (b) the medium dark "wash" represents the assay or flooding organism, always present when the assay organism displays growth; and (c) the light wash indicates inhibition of growth of the assay organism, inhibition zones, and these latter may overlap and thus eliminate any growth of the assay species on the plate.

FIGURE 13. Aerial hyphae of *Streptomyces aureofaciens* (A-377) at the time of spore differentiation.

FIGURE 14. OP-plate (OP) values of inhibition zones of *S. aureofaciens* (A 377-899) against *Bacillus cereus*.

FIGURE 15. Inhibition zones of *S. griseus* against *Staph. aureus*.

FIGURE 16. OP inhibition of *S. aureofaciens* against *Mycobacterium tuberculosis* (No. 607).

FIGURE 17. No zones of inhibition produced by *Penicillium chrysogenum* (Q 176) against *Mycob. tuberculosis* (No. 607).

FIGURE 18. Practically OP values of *S. aureofaciens* against *Coccobacillus* sp., an assay species strongly resistant to streptomycin.

FIGURE 19. OP inhibition of *S. aureofaciens* (A 377-899) against *Escherichia coli*.

FIGURE 20. Essentially no inhibition produced by *S. aureofaciens* against *P. chrysogenum* (Q 176).

FIGURE 21. No zones of inhibition in the test of *S. aureofaciens* (A 377-899) against *Aspergillus niger*.

FIGURE 22. Small zones of inhibition produced by *P. chrysogenum* (Q 176) against *S. aureofaciens* (A 377-899).

FIGURE 23. Practically OP values of *S. aureofaciens* (A 377-899) against *S. griseus* (A 1).

FIGURE 24. Zones of inhibition produced by *S. griseus* against *S. aureofaciens* (A 377-899).

THE PHARMACOLOGY OF DUOMYCIN

By B. K. HARNED, RAYMOND W. CUNNINGHAM, MARY C. CLARK,
RACHEL COSGROVE, CAROLYN H. HINE, WM. J. MCCAULEY,
EDGAR STOKEY, R. E. VESSEY, N. N. YUDA, AND Y. SUBBAROW

*Lederle Laboratories Division, American Cyanamid Company
Pearl River, N. Y.*

DUOMYCIN,* aureomycin A-377, is a yellow crystalline antibiotic obtained from the recently characterized mold *Streptomyces aureofaciens*.¹ Its potency is unusual both in degree and range in that it is effective against numerous gram negative and gram positive organisms as well as against some rickettsial and viral infections. This activity, coupled with a low toxicity, indicates that it is a potentially valuable chemotherapeutic agent worthy of extensive pharmacological analysis.

Animals and Methods

The acute toxicity of single doses has been studied in mice, rats, guinea pigs, rabbits, cats and dogs. The mortality count was made 72 hours after the dose was administered. The subacute toxicity from multiple doses has been studied in mice and dogs and the chronic toxicity in mice, rats, dogs and chickens. Including the pharmacodynamic studies, the problem has required 46 dogs, 24 cats, 36 rabbits, 109 guinea pigs, 559 rats, 760 mice and 18 chickens.

Electrocardiographic studies were made with a Cardiotron on anesthetized and unanesthetized dogs. Blood pressures were recorded in unanesthetized dogs by femoral arterial puncture and in anesthetized animals by cannulation of the carotid artery. Hemoglobin was determined as cyanmethemoglobin. The Wendel procedure² was used in the examination for methemoglobin. Clotting time was determined by the capillary tube method. Blood sugar was determined by the Shaffer-Somogyi method.³ The Lipschitz assay⁴ was used to evaluate diuretic activity.

Esbach's reagent was used in the tests for albuminuria. The filtered urines were treated in calibrated 15 cc. centrifuge tubes with the picric acid reagent, centrifuged, and compared with standardized serum albumin solutions treated similarly.

Phenolsulfonphthalein was used in the renal function tests in dogs. The routine was as follows: (a) Food was withheld from these animals for 24 hours; (b) fifteen cc. of water per kilogram of body weight was administered orally and, thirty minutes later, the bladder was drained and washed with 0.9 per cent sodium chloride; (c) the mono-sodium salt of phenolsulfonphthalein was injected intravenously in a dose of 0.4 mg.

* Trade mark

per kg.; (d) at one hour and at two hours, the dogs were catheterized and the bladders washed with 0.9 per cent sodium chloride; (e) the specimens were alkalinized and the amount of dye determined with a photoelectric colorimeter (filter No. 540). Normal dogs excreted from 60 to 85 per cent during the first hour.

Liver function was estimated by injecting intravenously 5 mg. per kg. of bromsulphalein. Blood was drawn at 5 and 30 minutes subsequent to the dose and the amount of the dye retained in the serum was estimated with a photoelectric colorimeter (filter No. 560). Normal retention at thirty minutes was from 0 to 15 per cent.

Tests for histaminic action were made on isolated guinea pig gut and also by following the changes in blood pressure in anesthetized cats. Antihistaminic action was tested on the isolated guinea pig gut and on guinea pigs in a spray chamber.

Tests for antipyretic action were made by following the rectal temperatures of rabbits injected intravenously with 0.25 cc. per kg. of typhoid vaccine, and of rats injected subcutaneously with 10 cc. per kg. of a 15 per cent suspension of yeast.⁵

Tests for irritation were made by the intracutaneous injection into the guinea pig skin and by local application to the conjunctival sac of the rabbit eye. In the eye, three or six drops at five minute intervals were used. Blood, urine and cerebrospinal fluid levels of duomycin were estimated biologically against *Bacillus subtilis* using a four hour period.⁶

Duomycin is soluble in acid and in alkaline solutions, but is almost insoluble at pH 7. Four per cent solutions may be easily prepared as the hydrochloride pH 2.5 or as a sodium salt in a carbonate buffer pH 8.5. At pH 2.5, the salt is stable but at pH 8.5, 25° C., it loses 12 per cent of its activity in 30 minutes, 20 per cent in 1 hour, and 40 per cent in 2 hours. Duomycin was administered parenterally usually as a 1 or 2 per cent solution pH 2.5 or pH 8.5. The hydrochloride was always used for the oral doses unless a different form was specified.

Acute Toxicity

Single Doses, Unanesthetized Animals. Orally, the toxicity is of a low order. Mice tolerated 1500 mg. per kg. and rats, 3000 mg. per kg. At 2500 mg. per kg., the mouse-mortality was 5 per cent (TABLE 1). No attempt was made to determine the maximal tolerated dose in larger animals. Intravenously, the L.D.₅₀ for the hydrochloride was 134 mg. per kg. for mice and 118 mg. per kg. for rats (FIGURE 1). The alkaline form appeared to be a little more toxic in mice but less toxic in rats (TABLE 1). These differences, however, are not significant. Three unanesthetized dogs readily tolerated intravenously, doses of 50 mg. per kg., pH 2.5,* given at a rate of 10 mg. per kg. per minute. Some days later, this dose in

* Doses larger than 30 mg. per kg. of duomycin hydrochloride pH 2.5 produce hemoglobinuria. An equivalent quantity of hydrochloric acid will produce the same result. Hemoglobinuria has never been observed after the injection of duomycin pH 8.5, even with doses of 100 mg. per kg.

TABLE 1 ACUTE TOXICITY OF DUOMYCIN

TABLE I. ACUTE TOXICITY OF DIOXIN

L.D.₅₀

19/20
Radial
limits
mg./kg.

Oral dose—mg./kg.		1000	1250	1500	1750	2000	2500	3000					
Mice	in 2% Starch pH 2.5	No. Injected No. Dead	20 0	20 0	20 0	20 1	20 1	20 7					
	in 2% Starch pH 2.5	No. Injected No. Dead	15 0	3 0	15 0	3 0	15 0	10 0					
Intravenous dose—mg./kg.*		40	50	60	70	80	90	100	125	140	150 175		
Mice	Aqueous pH 8.5	No. Injected No. Dead	11 0	14 1			11 4	10 9	5 5	102 88.7–117.3			
	Aqueous pH 2.5	No. Injected No. Dead	50 1	10 0	10 0	90 1	50 1	50 0	13 24	10 3	13 11 131 115.5–155.4		
Rats	Aqueous pH 2.5	No. Injected No. Dead	3 0	2 0	2 0	2 0	2 0	10 0	8 5	6 5	118 103.5–134.5		
	Aqueous pH 8.5	No. Injected No. Dead					10 1	6 1					
Guinea Pigs	Aqueous pH 2.5	No. Injected No. Dead					5 2	3 3					
	Aqueous pH 8.5	No. Injected No. Dead					5 2						
Subcutaneous dose—mg./kg.		100	200	300	400	500	600						
Mice	Aqueous pH 2.5	No. Injected No. Dead	8 0	8 0	4 0	10 2	10 3	10 1					
	Aqueous pH 8.5	No. Injected No. Dead	2 0	3 0									

* Injection Rate: Mice, 250 to 500 mg./kg./min.; rats, 35 mg./kg./min.; guinea pigs, 50 mg./kg./min.
Weight range: Mice—14-24 gms., rats—100-315 gms., guinea pigs—327-680 gms.

TABLE 1-A
THE EFFECTS OF MULTIPLE INTRAVENOUS DOSES OF DUOMYCIN PH 8.5 ON DOGS

Schedule: Blood counts and function tests on Feb. 20, 23, 24 and March 3 and 4. Duomycin doses Feb. 24, 10 mg./kg. A.M., 30 mg./kg. P.M.
Feb. 26, 26, 27, 20 mg./kg. A. M. & P.M., Feb. 28, 29, 20 mg./kg. A.M., March 1, 2, 20 mg./kg. A.M. and P.M.

*Bromsulphalein test
for liver function*

Dog No.	Weight		Time after injection of bromsulphalein		Clotting time of blood		Phenolsulfonphthalein test For kidney function	
	Feb. 20	Mar. 2	Apr. 4	Retention	Before Duomycin	After Duomycin	Before Duomycin	After Duomycin
	(kg.)	(kg.)	(kg.)	(per cent)	(min.)	(min.)	(per cent)	(per cent)
0260	7.6	7.1	7.4	16	9	21	11	71
0257	6.1	5.8	6.8	8	4	14	5	71
2669	7.9	7.3	8.3	28	10	22	6	73
2833	7.0	6.6	7.8	23	6	38	17	72
249	6.4	6.0	6.7	18	8	16	6	64
278	6.8	6.3	7.1	35	4	45	10	75

Blood Counts

Dog No.	Hemoglobin		Red blood cells		White blood cells		Neutrophils		Eosinophils		Lymphocytes	
	Before Duomycin	After Duomycin	Before Duomycin	After Duomycin	Before Duomycin	After Duomycin	Before Duomycin	After Duomycin	Before Duomycin	After Duomycin	Before Duomycin	After Duomycin
	(gms./100 cc.)	(gms./100 cc.)	(millions per cu. mm.)	(millions per cu. mm.)	(thousands per cu. mm.)	(thousands per cu. mm.)	(per cent)	(per cent)	(per cent)	(per cent)	(per cent)	(per cent)
0260*	17.9	15.8	8.13	8.80	12.0	11.1	51	55	15	6	33	38
0257*	13.9	12.6	7.45	6.70	10.9	14.7	48	65	10	7	41	28
2669*	13.9	15.2	7.41	7.60	7.2	20.5	44	68	6	2	50	29
2833*	14.0	15.8	5.47	8.35	15.1	7.8	45	51	12	3	40	46
249	16.5	15.4	8.05	7.59	5.6	19.9	50	51	4	4	46	49
278	16.5	14.9	7.09	7.45	13.95	27.2	50	70	10	4	40	26

* Dogs 0260 and 0257 had 1% of basophils before duomycin, 0260 and 2669 had 1% of monocytes after duomycin, and 2833 had 3% of monocytes before duomycin.

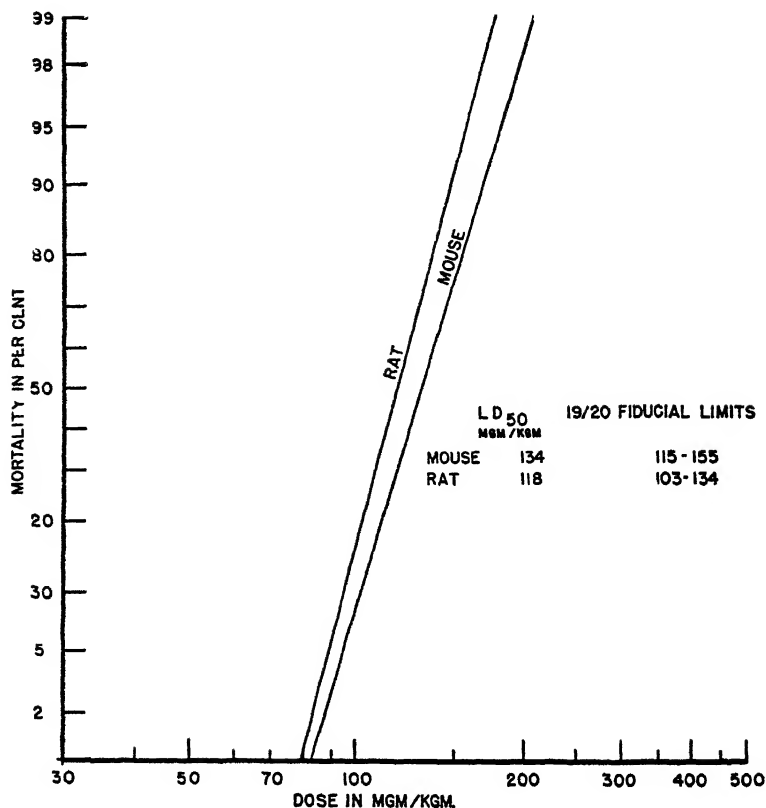


FIGURE 1. Acute intravenous toxicity of duomyen—pH 2.5. The experimental results were plotted on log probability paper and straight lines were fitted by eye. The 19/20 fiducial limits were estimated by a modification of the method of Litchfield and Fertig.⁹ The modification allowed for the fact that the population was not homogeneous in all cases and consisted of correcting the estimated values by multiplying by $\sqrt{(h+1)/n}$, as described by Wilcoxon & McCallan.¹⁰ The range of weights in grams was: mice 14–25; rats 100–315. Injection rates in mgm/kgm. min. were: mice 250–500; rats 30–40.

carbonate buffer, pH 8.5, was given to two of these dogs. No serious symptoms were observed. Two other dogs were given intravenously, three 20 mg. per kg. doses, and orally two 50 mg. per kg. doses in 24 hours. One dog vomited an oral dose but it was repeated later and retained. No other unfavorable symptoms were observed.

Three cats easily tolerated 50 mg. per kg., pH 8.5, injected intravenously at a rate of 50 mg. per kg. per minute. One cat vomited but other unfavorable reactions were not observed. Three rabbits and three guinea pigs were given intravenously 50 mg. per kg., pH 8.5, at a rate of 10–15 mg. per kg. per minute. There were no objectionable symptoms.

Subcutaneous, intramuscular and intraperitoneal doses at pH 2.5, 7.0 and 8.5 are irritating, and this factor plus the low solubility at the pH of the tissues renders the validity of toxicity data thus obtained open to

question. Mice and guinea pigs tolerated subcutaneous doses of 300 mg. per kg. (TABLE 1). Rats and mice tolerated 200 mg. per kg. intraperitoneally but this dose killed fifty per cent of the guinea pigs.

Tests for methemoglobin in mice, rats, guinea pigs, rabbits, cats and dogs were negative

Subacute Toxicity

Multiple Intravenous Doses to Unanesthetized Dogs. Over a period of eight days, six dogs were each given 270 mg. per kg. intravenously. All injections were made into the saphenous veins. Toward the end of the period, there was a definite amount of swelling near the sites of the injection, probably due to small amounts of the compound which entered the tissues around the veins. The dogs were lame and it is believed that the physical discomfort caused a loss of appetite which led to a decrease in weight (TABLE 1-A). It is interesting to note that one month after the administration of the compound, the weights of the dogs were equivalent to, or in excess of their predosing level. Their condition was excellent and although no biopsies were made the injected veins appeared normal.

Electrocardiograms were made before the drug and on the fourth day of the injections. No changes were noted. The data recorded in TABLE 1-A show that there were no changes in the complete blood counts, clotting time, liver function or kidney function.

Multiple Intramuscular Injections. Five guinea pigs were given 40 mg. per kg., pH 8.5 two times per day for four consecutive days and again on the sixth day. A second group of five was given half as much on the same schedule. There were no changes in hemoglobin or the total number of red blood cells in either group. Mice treated similarly had nine per cent reduction in hemoglobin and in red cells. Difference between the two doses were not significant. One mouse in each group, and one guinea pig on the 40 mg. schedule died.

Two dogs, No. 541 and No. 547, six months of age, were given eight 20 mg. per kg. doses of duomycin, pH 8.5 during a five day period. Dog 541 showed no change in weight, hemoglobin or red blood cells. Dog 547 lost ten per cent of his weight, thirteen per cent of his hemoglobin and twenty-one per cent of his red blood cells. Neither dog showed a change in the bromsulphalein test for liver function.

In all of the animals tested, intramuscular doses were irritating and produced an edema and tenderness in the injected area.

Chronic Toxicity of Oral Doses

Mice. Two groups of mice were given oral doses of duomycin pH 2.5. One group received 40 mg. per kg. and the other 100 mg. per kg. daily. At the end of the fourth week, the dose of the latter group was increased to 100 mg. per kg. two times a day. The growth curves for a twelve week period were the same for the treated groups as for the control group

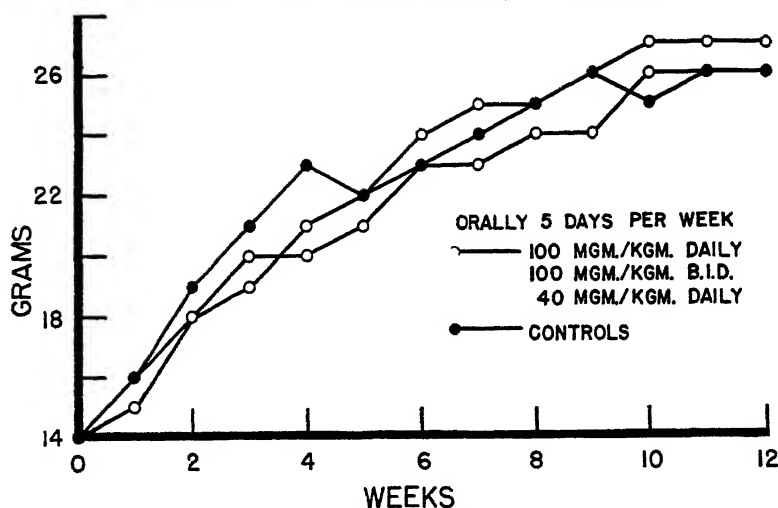


FIGURE 2. Growth curves on duomycin—male mice. After four weeks the dose of the group on 100 mgm/kgm was increased to 200 mgm/kgm.

Number of	Start	End
control	20	17
40 mgm/kgm	20	16
200 mgm/kgm	20	18

TABLE 2

EFFECTS OF MULTIPLE DOSES OF DUOMYCIN ON THE HEMATOLOGY OF RATS AND MICE

RATS AND MICE					Total blood cell count		
Animal	Number counted	Oral dose		Hemoglobin	Red blood cells	White blood cells	
		Per day* (mg. kg.)	Dura- tion** (weeks)		During final week on duomycin (gms. '100 cc.)	During final week on duomycin (millions per cu. mm.)	During final week on duomycin (thousands per cu. mm.)
Rat	8	200†	14	14.7	8.21	17.3	
Rat	8	40	14	14.6	9.26	17.04	
Rat	8	10	14	14.6	9.02	19.5	
Rat	8	0	14	14.3	8.39	21.4	
Mouse	10	200†	14	14.4	11.42	15.0	
Mouse	10	40	14	12.9	10.04	13.5	
Mouse	10	0	14	13.4	9.51	18.1	

* The daily dose was divided into two equal portions and administered in the morning and afternoon.

** Animals were dosed 5 days per week.

† For the first 17 doses the daily dose was 100 mg./kg.

(FIGURE 2). There were no unfavorable effects upon general appearance, the hemoglobin or the total blood cell counts (TABLE 2).

Rats. The data on rats are similar to those obtained on mice. The doses for the rat groups were identical with those used for mice except that an additional group was placed on 10 mg. per kg. per day (FIGURE 3). The

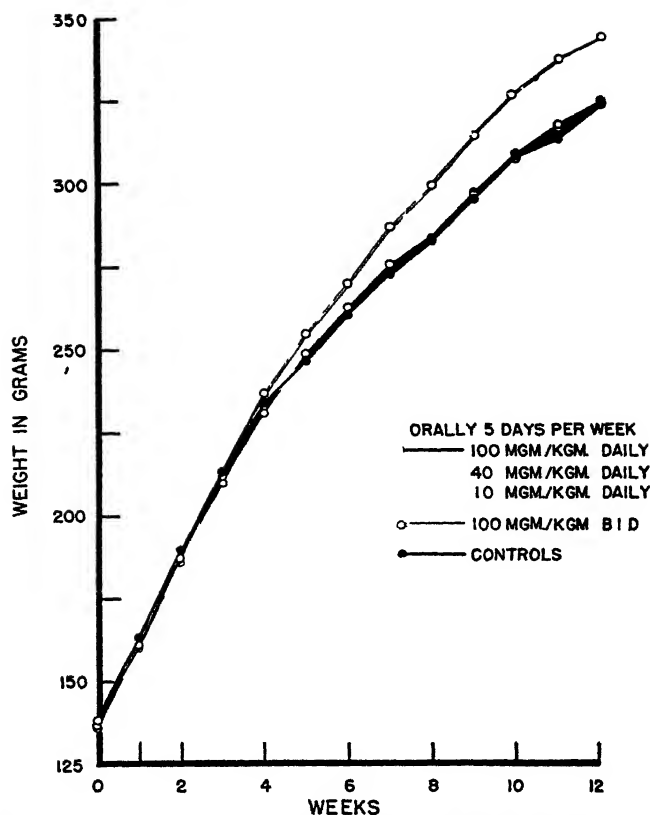


FIGURE 3. Growth curves on duomycin—male rats. After four weeks the dose of the group on 100 mgm./kgm. was increased to 200 mgm./kgm.

Number of rats	Start	End
control	20	18
10 mgm./kgm.	20	18
40 mgm./kgm.	20	20
200 mgm./kgm.	20	19

control group and the dosed groups grew at the same rate. The groups were indistinguishable and there was no effect upon the hemoglobin or blood cells (TABLE 2). At the end of the period, the blood pressure of these animals was tested by Mr. Ablondi with the method and apparatus recently developed in his laboratory.⁷ There was no significant difference between the mean pressures of the groups. The blood sugars of the dosed and control groups were the same.

*Pathologist's Report.** Five mice from the control group, 10 from the 40 mg. per kg. group and 10 from the 100–200 mg. per kg. group were sacrificed. Similar numbers from the four groups of rats were also sacri-

* The pathological examinations were made by E. Woll, M. D.

TABLE 3
THE EFFECTS OF MULTIPLE ORAL DOSES OF DUOMYCIN ON DOGS
*Bromsulphalein test
for liver function***

Dog No.	Oral dose		Weight		Time after injection of bromsulphalein			During final week		Totaling time of blood†
	Per day*	Dura- tion** (weeks)	Before duomycin (kg.)	Final day on duomycin (kg.)	5 min.	30 min.	5 min.	Retention (per cent)	Before duomycin (min.)	
500	40	1								
	100	15	2.2	4.6			5	18	4.5	3.2
523	40	1								
	100	15	3.9	4.6			3	34	2.2	3.2
491	40	1								
	100	15	2.7	7.2			6	53	2.5	7.0
548	100	12	4.6	9.1			6	39	4.3	8.5
589	100	10	8.1	9.9	22		8	20	9.5	11.2
136	100	9	8.1	7.4	24		5	—	11.5	5.2
263	100	10	8.4	9.5	13		3	23	12.0	3.3
587	100	9	7.2	8.0	23		12	21	3.2	2.5
586	100	10	9.5	9.9	24		3	45	7.0	6.5
317	100	10	8.3	8.8	38		12	27	9.5	11.2

TABLE 3 (continued)

Dog No.	Phenolsulfonphthalein test for kidney function**		Albuminuria** Esbach's test during final week on diuomycin (per cent)	Blood pressure†† during final week on diuomycin (mm./Hg)
	Before diuomycin	During final week on diuomycin		
	Two hour excretion (per cent)			
500	—	—	—	124
523	—	—	0	130
491	—	—	0	124
548	—	—	0	106
589	—	—	0	122
136	75	74	<0.025	110
268	85	66	0	144
587	76	68	0	114
586	80	81	0	124
317	85	84	0.025	116

* The daily dose was divided into two equal portions and administered in the morning and afternoon.

** Dogs were not dosed on Saturdays or Sundays. Tests for function were not made on Mondays.

† Capillary tube method.

†† From puncture of femoral artery.

TABLE 4
THE EFFECTS OF MULTIPLE ORAL DOSES OF DUOMYCIN ON THE HEMATOLOGY OF DOGS

Dog No.*	Hemoglobin		Total blood cell count				Differential count				(per cent)	
	Before week on duomycin	During final duomycin	Red blood cells		White blood cells		Neutrophils		Eosinophils			Lymphocytes
			Before duomycin	During final week on duomycin	Before duomycin	During final week on duomycin	Before duomycin	During final week on duomycin	Before duomycin	During final week on duomycin		
	(Gms./100 cc.)		(millions per cu.mm.)	(thousands per cu.mm.)	(per cent)	(per cent)	(per cent)	(per cent)	(per cent)	(per cent)	(per cent)	
500	11.4	13.9	5.30	7.00	19.3	9.7	69	—	0	—	31	
523	11.6	17.4	4.85	8.85	23.8	8.2	75	—	2	—	23	
491	11.4	13.9	5.90	7.15	17.2	12.2	46	—	4	—	50	
548	10.2	13.1	5.10	7.50	—	10.0	47	—	0	—	53	
589	9.6	10.5	5.15	5.90	9.3	12.6	71	—	1	14	34	
136	12.2	13.6	5.45	6.70	21.6	21.8	53	53	30	24	5	
268	17.2	14.5	8.50	6.20	14.9	10.4	68	46	4	12	28	
587	15.6	12.1	7.50	6.10	10.3	13.8	49	74	5	6	46	
586	14.5	16.8	7.65	7.95	11.0	14.8	84	72	2	14	14	
317	14.9	15.6	7.50	6.50	8.1	6.8	50	49	12	18	39	

* For doses see TABLE 3.

ficed. "No morphological changes attributable to duomycin were seen on gross or microscopic examination. In the mice and rats, there was some low-grade bronchitis. However, the incidence of spontaneous pulmonary pathology so common in this species was decidedly less in the treated groups."

Dogs. Ten dogs were given daily doses of 100 mg. per kg. orally in capsules for periods of 9 to 15 weeks. The essential data shown in TABLES 3 and 4 demonstrate that duomycin did not change the bromsulphalein test for liver function, the P.S.P. test for kidney function, the hematology or the clotting time. Seven of the nine dogs showed no albuminuria. Of the others, one had 0.025 per cent albumin and one less than 0.025 per cent. Through error, no test was made on dog No. 500. Tests for albuminuria were not made prior to dosing, hence it is not known if the traces found in the two dogs were present before duomycin. Blood pressures taken during the final weeks on duomycin were within the normal range. At the end of the experimental period these animals were in excellent condition.

*Pathologist's Report.** Dogs 587, 136, 523, 548, 589, and 491 were sacrificed and studied. Intestinal parasites were found in 589, 548, and 523. Heart worms were found in 136, nephrosclerosis in 587 and 136 and simple ovarian cysts in 587. The pathologist concluded "Duomycin administered orally in the above schedule did not produce any detectable gross or microscopic changes."

Chicks. Robbins' observation⁸ that 2,4-dinitrophenol produced cataracts in the eyes of chicks prompted us to subject duomycin to similar test. Eighteen seven-day-old chicks were divided into groups of six and given the following compounds in the chick feed: Duomycin hydrochloride, 0.25 per cent; 2,4-dinitrophenol, 0.25 per cent. Plain chick feed was given to the controls. The chicks on 2,4-dinitrophenol developed cataracts overnight. The chicks on duomycin showed no lenticular changes after seven days of dosing. No chick in this group died. Two were lost from the control group. It is interesting to note that, although the average weight of the chicks in the two groups was the same at the start, at the end the birds on duomycin had an average weight of 99 grams as compared with 81 grams for the controls. Furthermore, the dosed chicks appeared to be more healthy than the controls and after 6 weeks this superiority was maintained. Apparently, the duomycin had eliminated some infection.

Tests for Pharmacological Activity

Blood Pressure. DOGS. The data in TABLE 5 and FIGURES 4, 5 and 6 show the effects on the blood pressure of small and large doses at pH 2.5 and 8.5 given to etherized and nembutalized dogs at different rates of injection. FIGURE 4 records an observation duplicated many times. Under the same conditions, the rapid injection of duomycin hydrochloride

* The pathological examinations were made by E. Woll, M. D.

TABLE 5
THE EFFECTS OF MULTIPLE INTRAVENOUS DOSES OF DUCOMYIN ON THE BLOOD PRESSURE OF ANESTHETIZED ANIMALS

Animal	Anesthetic	Number	Dose		Time required for infection (seconds)	Interval between injections (min.)	Blood pressure			
			(pH)	(mg./kg.)			Control (mm. Hg)	Full Maximum (mm. Hg)	Rise Maximum (mm. Hg)	Stabilized at (mm. Hg)
Dog 611	Ether	1 2, 3, 4, 5, 6, 7 8, 9, 10	8.5	10	25	15	120	6	4	124
Dog 607	Ether	1 2, 3 4, 5 6, 7 8, 9	2.5 2.5 8.5 2.5 2.5	10 10 10 10 10	20 20 20 20 20	15 15 22, 13 12, 29 13, 18	106 132 150, 152 151, 156 160, 158	6 36, 30 3 4, 8 16	28 0.18 2 2.0 0	124 128 132 132-150 152, 154 156* 158, 156
Dog 608	Ether	1 2, 3, 4, 5 6, 7, 8 9, 10, 11, 12 13 11	2.5 2.5 2.5 2.5 2.5	10 10 10 10 10	20 20 20 20 20	15 15 15 15 15	160 194-190 186-171 170-154 150 146	18 10-22 16-24 18-38 38 40	34 0.6 0 0 0 0	194 194-186 180-170 164-150 146 134
Cat 619	Ether	1, 2, 3, 4 5 6	8.5†† 8.5†† 8.5††	10 10 50	8 30 60	10-15 10 15	104 106 100	0 0 0	0-2 4 22**	104-106 100 94
Cat 620	Ether	1 2, 3, 4 5, 6, 7 8, 9, 10, 11 12 13	8.5†† 8.5†† 8.5†† 8.5†† 8.5††	20 20 20 20 50	12 12 12 12 10	10 10 10 10 10	142 140-132 122-104 98-88 86 80	4 0 0 0 12	4** 6-8** 8-6** 6-12** 14** 24**	140 138-122 122-98 90-86 80 70
Cat 618	Ether	1 2, 3, 4, 5 6 7	8.5†† 8.5†† 8.5†† 8.5††	10 10 20 40	60 10 10 40	10 10 10 10	76 112-110 112 112	0 6 6 4	40** 4-8** 8** 12**	112 108-112 112 112

TABLE 3 (continued)

Animal	Anesthetic	Number	(pD)†	(mg./kg.) (seconds)	Time required for injection (min.)	Blood pressure			
						Control (mm. Hg)	Fall†† Maximum (mm. Hg)	Rise‡ Maximum (mm. Hg)	Stabilized at (mm. Hg)
Cat 661	Dial***	1, 2, 3, 4, 5 6, 7, 8, 9, 10, 11	8.5 8.5	5 5	20 20	120-128 130-120	0 0-3	0-3	120-130 130-120
Cat 662	Dial***	1, 2 3, 4, 5, 6, 7 8, 9, 10, 11	8.5 8.5 8.5	5 5 5	20 20 20	100 102-136 146	0 0 0	0-2 1-16 0	100-102 106-146 146
Cat 652	Dial***	1, 2 3, 4, 5, 6, 7 8	2.5 2.5 2.5	5 5 5	20 20 20	132, 136 140-150 150	10, 12 8-12 4	4 0 0	136, 140 140-150 150
Cat 654	Dial***	1 2, 3, 4, 5, 6	2.5 2.5	5 5	18 18	102 112-108	14 12-20	10 0	112 106-104
Cat 657	Dial***	1 2 3 4 5 6 7 8	2.5 2.5 2.5 2.5 2.5 2.5 2.5 2.5	10 10 10 25 10 10 10 25	10 30 60 300 10 30 60 300	136 152 144 162 136 142 146 158	46 22 6 4 136 32 28 6	24** 16 18 0 6 8 6 0	138* 168* 162 158* 142 150* 152* 158
Cat 655	Dial***	1, 2, 3 4, 5, 6, 7, 8 9, 10, 11	2.5 2.5	5 5	20 20	108-118 118-142	0 10-28	2-5 2-8	110-118 130-150
Cat 653	Dial***	1, 2, 3, 4, 5 6, 7	2.5 2.5	5 2.5	20 20	182-174 168-166	14-18 8	0 0	176-168 166

* When the stabilized level did not agree with the subsequent control value some other procedure had been interposed.

** Rise was followed by a gradual fall.

*** Dial-urethane. Each cc. contained: diallylbarbituric acid 0.1 gm.; urethane 0.4 gm.; monoethylurea 0.4 gm. The dose was 0.7 per cc. per kg., intraperitoneally.

† The duration of the fall of conyngin was 1.0 per cent unless stated otherwise.

†† The rise usually lasted two to four minutes.

‡‡ The concentration of duomycin was 2.0 per cent.

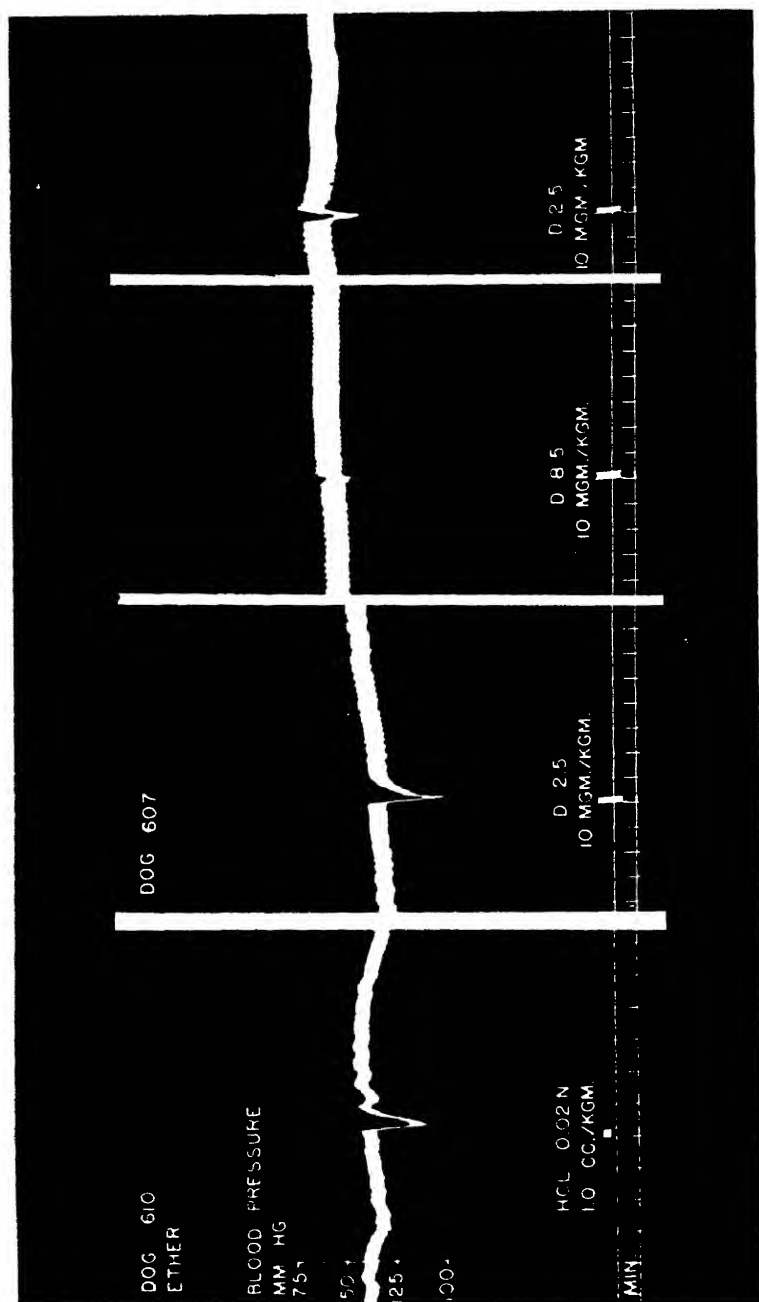


FIGURE 4. A comparison of the effects of an initial injection of diumycin hydrochloride pH 2.3 with an initial injection of hydrochloric acid equivalent in total acidity and volume. These are compared with the action of diumycin in a sodium carbonate buffer pH 8.5. All injections were intravenous.

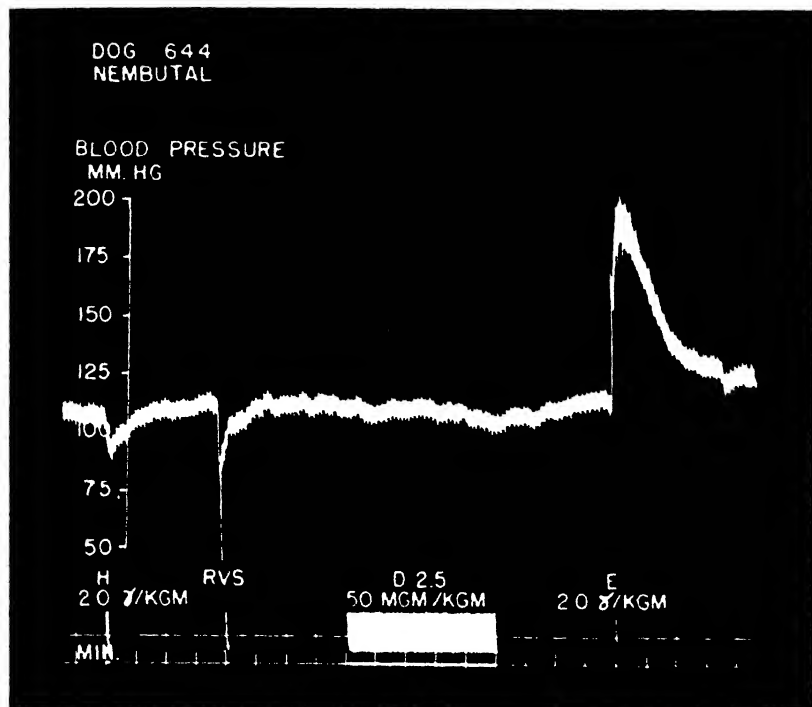


FIGURE 5. The effect of duomycin hydrochloride given intravenously at a rate of 10 mgm/kgm/min. The response to histamine diphosphate 2.0γ/kgm (H); stimulation of the right vagus (RVS); and to epinephrine 2.0γ/kgm (E) indicate that the vasomotor responses were normal.

ride at pH 2.5 produced a fall in blood pressure which is almost identical with that produced by an injection of hydrochloric acid equivalent in normality and total volume. In this figure, different dogs were used to permit comparison of initial doses. The results on dog 607 demonstrate also that the same dose and rate of injection of the alkaline duomycin affects the blood pressure less than the acid form.

The influence of the rate of injection may be seen by a comparison of the results shown in FIGURES 4 and 5. In dog 607 (FIGURE 4), the rate of injection was 30 mg. per kg. per minute and the blood pressure fell 31 mm. of mercury. In FIGURE 5, dog 644, the rate was 10 mg. per kg. per minute and no fall occurred even though the total dose was five times as great. In FIGURE 6, the rate of administration was 20 mg. per kg. per minute of the alkaline duomycin and the first injection produced a fall of 20 mm. of mercury, but the second injection of the same dose at the same rate produced no change in the blood pressure. The vasodepressor action of this rate appeared to be borderline.

The total dose tolerated in a relatively short period is high. Dog No. 611, TABLE 5, was given 100 mg. per kg., pH 8.5 in 140 minutes. After

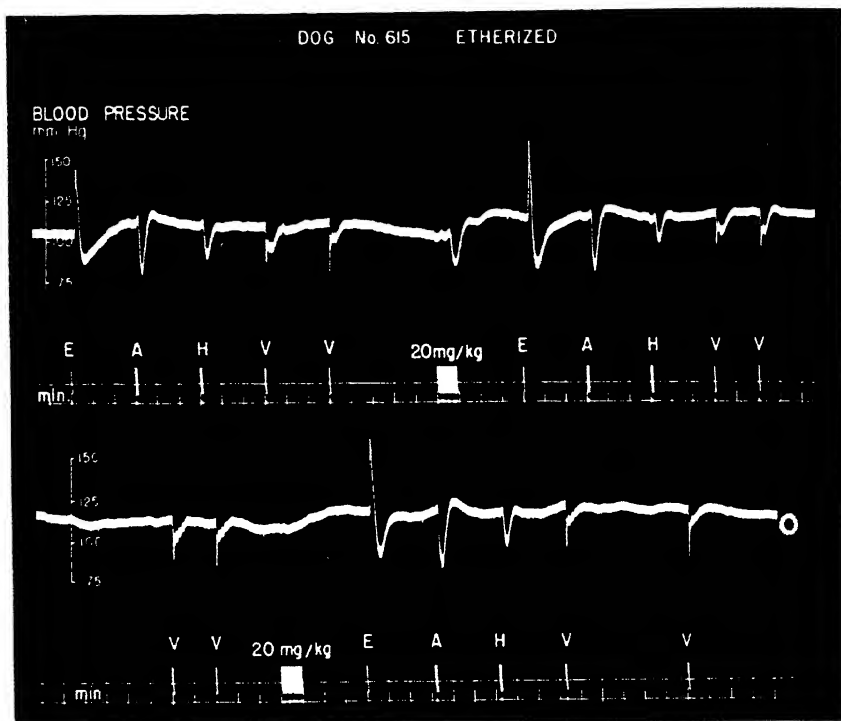


FIGURE 8. The effect on the blood pressure of the intravenous injection of duomycin pH 8.5 at the rate of 20 mgm/kgm min. The vasodepressor action of this rate appears to be border line. Duomycin did not modify the vasomotor responses of: epinephrine, 2.0, kgm (E); acetylcholine chloride, 0.4, kgm (A); histamine diphosphate, 2.0, kgm (H); stimulation of the right vagus (V). See also TABLE 6.

the last injection, the blood pressure was 128 mm. of mercury, 8 mm. above the control value. Dog 608 was given 140 mg. per kg., pH 2.5, in 215 minutes. The blood pressure before duomycin was 160 mm. of mercury and after the last dose it was 134. In dog 607, the first injection of duomycin pH 2.5 raised the blood pressure from 106 to 132. After eight more injections of 10 mg. per kg. each during a period of 110 minutes, the pressure was 156 mm. of mercury. Dog No. 665 anesthetized with nembutal was injected intravenously with the alkaline duomycin pH 8.5 at a rate of 6.25 mg. per kg. per minute. During a period of forty minutes, a total dose of 250 mg. per kg. was given. The control blood pressure of 143 mm. of mercury remained essentially unchanged and at the end of the injection the pressure was 146.

In 4 unanesthetized dogs, the intravenous injection of 20 mg. per kg. in thirty seconds produced no important changes in the blood pressure (FIGURE 9). The dog with the greatest change was a poorly trained animal with an initial blood pressure of 140, a pressure just before the injection of

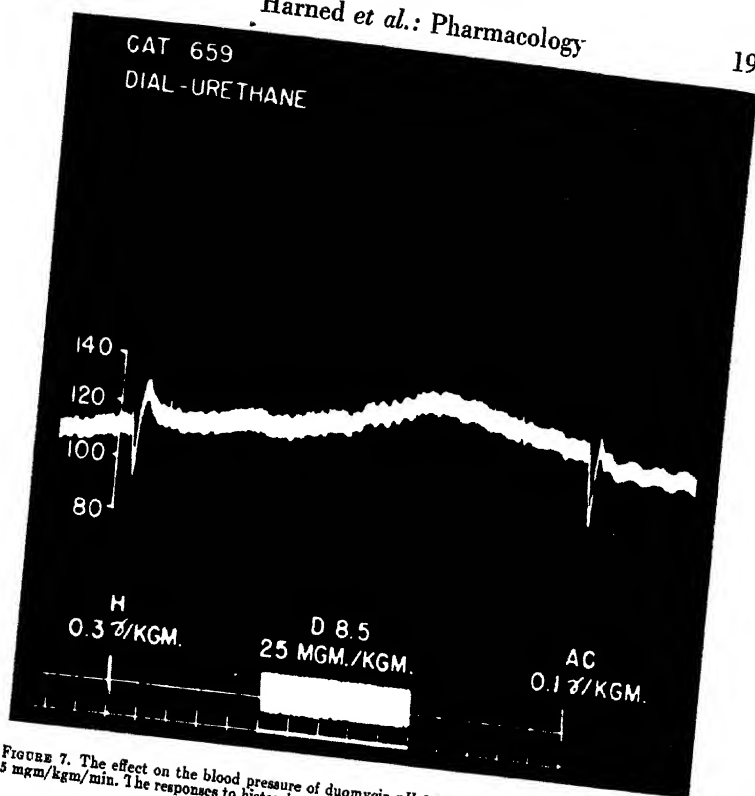


FIGURE 7. The effect on the blood pressure of duomycin pH 8.5 injected intravenously at the rate of 5 mgm/kgm/min. The responses to histamine base (H) and to acetylcholine chloride (AC) were normal.

128 and a final pressure of 118. The excitement incident to the preparation of the dog may account for the higher control value.

CATS. The blood pressure data on cats anesthetized with dial-urethane or with ether present a picture almost identical with that obtained on dogs (TABLE 5, FIGURES 7 and 8). Perhaps attention should be called to cat 646 (FIGURE 8). This animal was given 100 mg. per kg. of duomycin pH 2.5 at a rate of 12.8 mg. per kg. per minute. During the first few seconds of the injection the blood pressure fell from 170 to 120 and rose immediately to 165 mm. of mercury. This effect appears to be associated with the acidity of the solution, since a similar effect was observed after the injection of hydrochloric acid. After 35 to 40 mg. per kg. of duomycin had been injected, the pressure fell gradually to 85 mm. of mercury, but rose to 130 and stabilized as soon as injection had ceased.

Effects of Duomycin on Vasomotor Responses. In either the acid or alkaline form, duomycin does not modify the response to epinephrine, histamine, acetylcholine or faradic stimulation of the right vagus (FIGURES 6 and 8 and TABLE 6). The tests were made on dogs and cats anesthetized

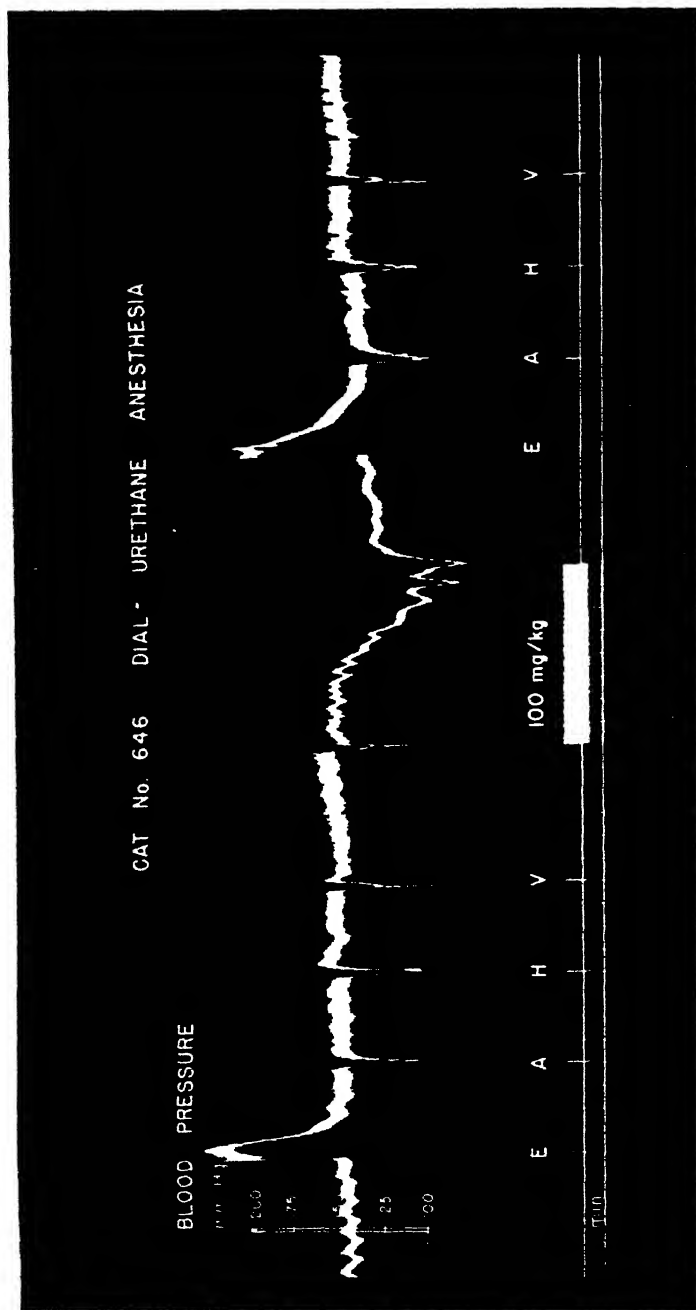
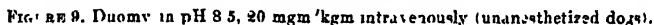


FIGURE 8. The effect on the blood pressure of the intravenous injection of a massive dose of duomycin pH 2.5 administered at the rate of 12.8 mgm/kgm/min. Duomycin did not modify the vasomotor responses of epinephrine, 4.07/kgm (E); acetylcholine chloride, 0.17/kgm (A); histamine diphosphate, 0.87/kgm (H); stimulation of the right vagus (V). See also TABLE 8.



In TABLE 7, the results from some intravenous doses in dogs and rabbits

TABLE 6

THE RESPONSE OF THE BLOOD PRESSURE TO EPINEPHRINE, ACETYLCOLINE, HISTAMINE, AND FARADIC STIMULATION OF THE RIGHT VAGUS BEFORE AND AFTER INTRAVENOUS ADMINISTRATION OF DUOMYCIN

Animal	Anesthetic	Dose of duomycin† (pH)	Dose (mg./kg.)	Blood pressure		Blood pressure		Interval between duomycin and epinephrine (min.)	Acetylcholine-fall		Interval between duomycin and acetylcholine (min.)		
				Epinephrine-rise		Control			Before After				
				Before (mm. Hg)	After Duomycin (mm. Hg)	Before Duomycin (mm. Hg)	After Duomycin (mm. Hg)		Before Duomycin (mm. Hg)	After Duomycin (mm. Hg)			
Dog 614	Ether	8.5	20	150	148	14	12	3	156	146	36	42	6
Dog 614	Ether		20		140		42	3		150		40	6
Dog 615	Ether	8.5	20	102	112	38	14	3	108	114	28	30	6
Dog 615	Ether		20		116		42	3		116		32	6
Cat 620.	Ether	8.5*	60**	122	102	22	12**	5	124	96	24	18**	10
Cat 618	Ether	8.5*	110**	76	118	14	22**	25	90	112	34	32**	
Cat 620	Ether	8.5*	290**	142	70	16	22**	5	142	66	20	14**	
Cat 649	Dial***	8.5	50	160	134	34	48	5	162	142	20	14	10
Cat 659	Dial***	8.5	25		132	36	42	3	106	122	20	26	5
Cat 650	Dial***	8.5	50	142	126		36	3	136	130	38	30	6
Cat 658	Dial***		25***						124		28		6
Cat 658	Dial***	8.5	25						120	104	22	22	5
Cat 658	Dial***		25***						110	112	26	24	5
Cat 658	Dial***		25***						116	116	0	0	5
Cat 658	Dial***		25***						110	126	8	8	8
Dog 644	Nembutal, Sodium	2.5	50	100	112	80	80	4	120	120		12	34
Dog 644			100		124		80	30	112	96	26	14	10
Cat 647	Dial***	2.5	50	108	94	68	64	5	112	96	30	46	14
Cat 646	Dial***	2.5	50	170	148	54	62	10	162	150	30	46	14
Cat 646	Dial***		100		134		64	5	136	136	38	38	9

TABLE 6 (continued)

Animal	Blood pressure				Interval between duomycin and histamine (min.)	Blood pressure				Interval between duomycin and stimulation of vagus (min.)
	Control		Histamine-fall			Control		Vagal-fall		
	Before Duomycin (mm. Hg)	After Duomycin (mm. Hg)	Before Duomycin (mm. Hg)	After Duomycin (mm. Hg)		Before Duomycin (mm. Hg)	After Duomycin (mm. Hg)	Before Duomycin (mm. Hg)	After Duomycin (mm. Hg)	
Dog 614	150	144	20	28	9	142	146	34	96	12
Dog 614		148		26	9		142	34†	34	12
Dog 615	108	110	18	12	9	110	112	38	26	12
Dog 615		112		18	9		118	30†	26	12
Cat 626						122	96	16	10**	1.5
Cat 618										
Cat 620										
Cat 649	162	146	32	32	15	158	146	10	26	20
Cat 659	110	114	16	18	10					
Cat 650	134	134	36	36	9	138	136	56	56	12
Cat 650		124		34	9		124		48	12
Cat 658	116	118	14	22	10					
Cat 658	108	116	0	0	10					
Cat 658	116	120	0	0						
Dog 644	106	122	14	11	12	112	122	68	76	16
Dog 644		116		8	38		111		60	12
Cat 647	106	90	16	8	15	92	82	31	40	20
Cat 646	160	148	32	42	18	154	152	66	58	22
Cat 646		140		31	13		148		50	17

† A 1.0 per cent solution was used except in cats 618, 650, 626. The doses were not cumulative unless so stated.

* Concentration of solution—2.0 per cent.

** Three injections of 20 mg./kg. of duomycin at intervals of 10 minutes. Epinephrine, acetylcholine, histamine and vagal stimulation followed the last injection of duomycin.

†† Five injections of 10 mg./kg. of duomycin; one 20 mg./kg., and one 40 mg./kg. at intervals of 10 minutes. Epinephrine, acetylcholine, histamine and vagal stimulation followed the 40 mg./kg. of duomycin.

‡ Twelve injections of 20 mg./kg. of duomycin and one 50 mg./kg. at intervals of 10 minutes. Epinephrine, acetylcholine, histamine and vagal stimulation followed the 50 mg./kg. dose.

‡‡ After 2.0 mg. dose, intravenously, of Tagathen (an antihistaminic drug).

++++ After Tagathen and 0.1 mg./kg. atropine sulfate, intravenously.

††† New control.

*** Dial-urethane. Each cc. contained: diallylbarbituric acid 0.1 gm., urethane 0.4 gm.; monochlorefure 0.4 gm. The dose was 0.7 cc. per kg. intraperitoneally.

TABLE 7
 TESTS FOR ALBUMINURIA IN DOGS AND RABBITS DOSED INTRAVENOUSLY WITH DIOMYCIN
Time of collection of urines in hours subsequent to dose

Dog No.	Intravenous dose (mg./kg.) (pH)	Control sample	0.5	1	2	3	4	5	24	30	48	72	96	102
545	100 2.5	Trace <0.025							Trace <0.025	0			0	Trace <0.025
543	100 2.5	Trace <0.025	0.1		0.15				0.1	0				
542	50 8.5	0		0.1			0.1	Trace <0.025					Trace <0.025	0
556	50 8.5	0		0.1			0	Trace <0.025					0	0
Rabbit 666	50* 8.5	0							Trace <0.025					
Rabbit 667	50* 8.5	0									0			

* This dose was divided into two equal portions one of which was administered in the morning, the other four hours later

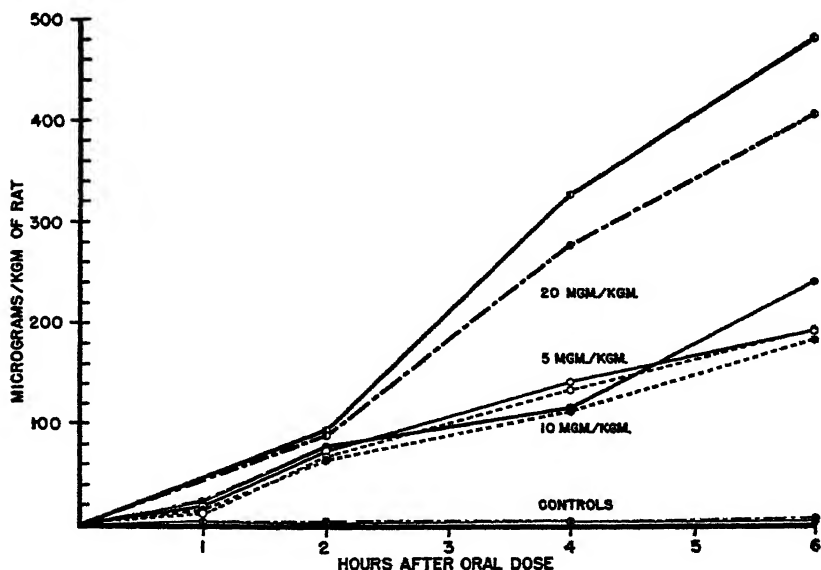
TABLE 8

TESTS FOR ALBUMINURIA IN RATS DOSED ORALLY WITH DUOMYCIN

Animal	No. of groups	No. of rats (per group)	Dose of duomycin		Collection period (hours)	Albuminuria (per cent)
			(mg./kg.)	(pH)		
Rats	12	8	0	—	0-5	Trace (<0.025)
Rats	2	8	50	2.5	0-5	Trace (<0.025)
Rats	4	8	100	2.5	0-5	Trace (<0.025)
Rats	3	8	200	2.5	0-5	Trace (<0.025)
Rats	1	8	400	2.5	0-5	Trace (<0.025)

have been recorded. Following doses of 50 to 100 mg. per kg. of duomycin, traces of albumin appeared temporarily in the urine of the dogs. Of the two rabbits given 50 mg. per kg., one showed a trace of albumin, less than 0.025, while the urine from the other was negative.

Urinary Excretion of Duomycin. One hour after the oral administration of duomycin hydrochloride to rats, the urine contained measurable quantities, even after doses as low as 5 mg. per kg. In FIGURE 10, it may be seen that the rate of excretion is relatively constant for at least six hours. Although the experiments were usually terminated in this period,



	TUBE DILUTION ASSAY				TURBIDIMETRIC ASSAY			
DOSE—MG./KGM.	0	5	10	20	0	5	10	20
NO OF EXPERIMENTS	6	5	4	3	4	4	2	2

FIGURE 10. Urinary excretion of duomycin (rats) Solid lines represent turbidimetric assay; the broken lines the tube dilution assay. Each experiment represents the urine from a group of four rats.

TABLE 9
DUOMYCIN- TESTS FOR IRRITATION IN THE RABBIT'S EYE

Rabbit No.	Duomycin preparation	pH	Observations		
			Time in hours subsequent to final drop		
			1 hour	24 hours	48 hours
1	0.5% in 0.9% NaCl†	2.9	mild irritation**	no irritation	no irritation
14*	0.5% in 0.9% NaCl†		mild irritation**	no irritation	no irritation
2	0.5% in 0.9% NaCl†		mild irritation**	no irritation	no irritation
15*	0.5% in 0.9% NaCl†		mild irritation**	no irritation	no irritation
3	0.5% in dist. H ₂ O	2.7	irritation	mild irritation	no irritation
16*	0.5% in dist. H ₂ O		mild irritation	mild irritation	no irritation
5	0.5% in dist. H ₂ O		mild irritation	mild irritation	no irritation
17*	0.5% in dist. H ₂ O		mild irritation	mild irritation	no irritation
4	0.9% NaCl	6.8	no irritation	no irritation	no irritation
6	0.9% NaCl		no irritation	no irritation	no irritation
11	0.9% NaCl		no irritation	no irritation	no irritation
7	1.0% in 0.9% NaCl	2.9	mild irritation	no irritation	no irritation
9	1.0% in 0.9% NaCl		mild irritation	mild irritation	no irritation
8	1.0% in dist. H ₂ O	2.5	mild irritation	mild irritation	—
13*	1.0% in dist. H ₂ O		irritation	mild irritation	—
10	1.0% in dist. H ₂ O		mild irritation	mild irritation	—
19*	1.0% in dist. H ₂ O		irritation	mild irritation	—
12	0.25% in 0.9% NaCl	2.9	mild irritation	no irritation	—
13	0.25% in dist. H ₂ O	2.8	no irritation	no irritation	—
20*	0.02N HCl	1.7	mild irritation	no irritation	—
21*	0.02N HCl		mild irritation	no irritation	—

Method used. One drop of the preparation was added to the left eye at five minute intervals until eight drops had been given.

* One drop of the preparation was added to the right eye at five minute intervals until three drops had been given.

** Mild irritation was manifested in temporary reddening of the conjunctiva.

† The borate complex appeared to be less irritating than the hydrochloride

one experiment showed that significant amounts were excreted between the tenth and twelfth hour. The collection-vessels were kept in iced water to limit the decomposition of the drug. During the first six hours, 3.8 per cent of the administered dose of 5 mg. per kg. was excreted. At doses of 10 and 20 mg. per kg., two per cent of the dose was excreted in 6 hours.

After intravenous injections, the excretion in dogs is rapid for the first two hours, somewhat slower between the second and fourth hours, and low between the fourth and sixth hours (FIGURE 11). This figure shows a similar quantitative excretion after the injection of duomycin either pH 2.5 or 8.5.

Blood Sugar. Food was withdrawn from rats 16 hours before they were given 150 mg. per kg. of duomycin hydrochloride orally. Bloods were taken at 1.5, 3, and 5 hours after dosing and blood sugar determinations

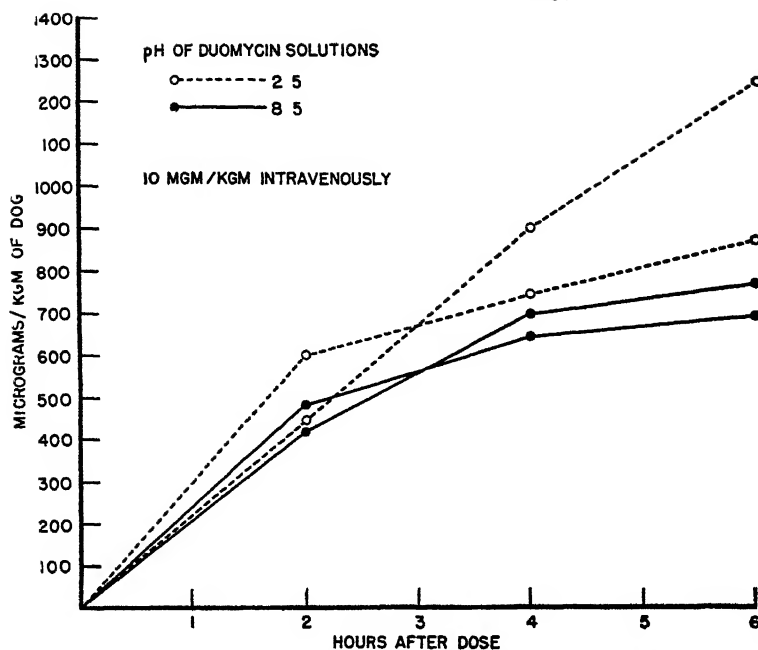


FIGURE 11. A comparison of the excretion of duomycin injected intravenously as the hydrochloride pH 4.5 and in a sodium carbonate buffer pH 8.5. Two dogs were injected with each solution.

were made on zinc hydroxide filtrate by the Shaffer-Somogyi method.⁴ The average fasting blood sugar was 71 mg. per cent and there was no significant deviation from this value during the period of the test. Blood sugar determinations made on the dogs and rats used in the chronic dosing experiments were also normal.

Antihistaminic Action. Ten guinea pigs given intravenously 25 mg. per kg. 30 minutes before testing showed neither protection nor potentiation of the effects of the standard histamine spray.

Action on Isolated Tissues. Duomycin had essentially no effect on the isolated gut from the guinea pig or rabbit. Tests for activity on the isolated uterus from these species were also negative.

Antipyretic Action. No evidence of antipyretic activity was observed in fevered rabbits and rats. FIGURE 8 is a composite of four experiments made on different days. The conclusion from each individual experiment was the same.

Irritation. Duomycin pH 2.5, 7, or 8.5 is irritating when injected intraperitoneally, intramuscularly, subcutaneously, or intracutaneously. In the rabbit eye, a 0.5 per cent solution of duomycin hydrochloride in 0.9 per cent saline is mildly irritating for a short time subsequent to the administration, but 24 hours later no evidence of irritation can be detected. The results are shown in TABLE 9. The mildness of the irritation

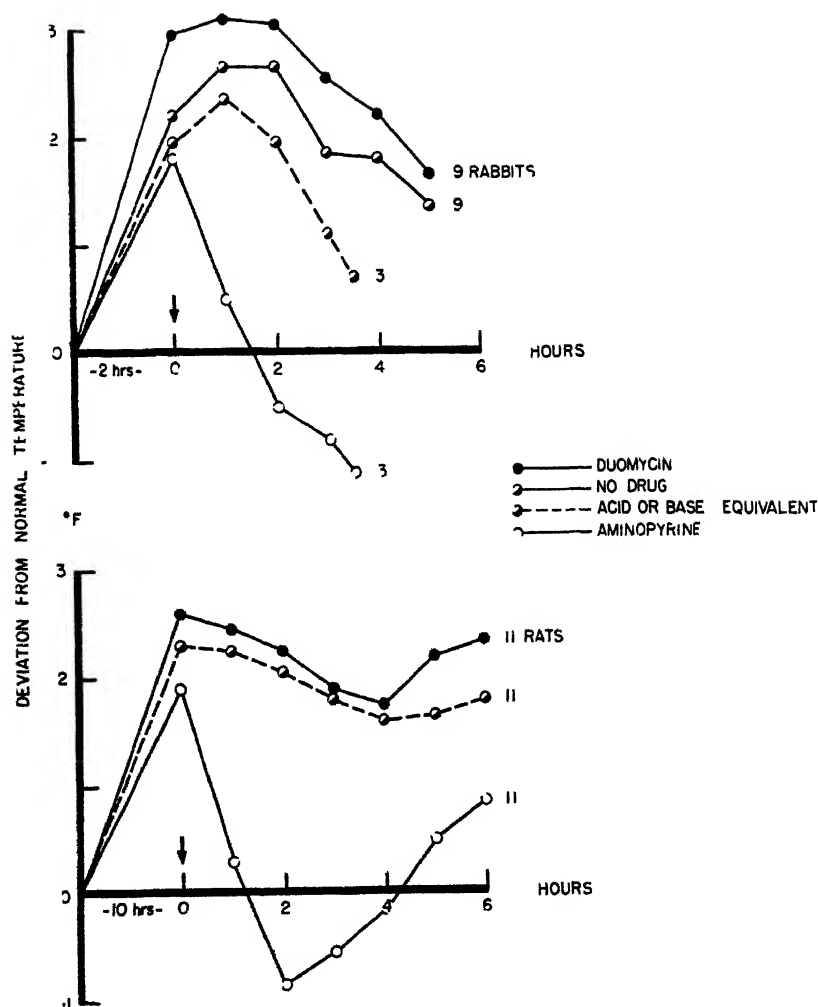


FIGURE 12. Antipyretic action of duomycin in rabbits and rats.

Fever was produced in rabbits by the intravenous injection of 0.25 cc. per kg. of combined typhoid vaccine. Two to two and one half hours were allowed for the development of the fever before the test-compounds were administered. At the time indicated by the arrows, duomycin pH 8.3 was injected intravenously in a dose of 25 mg. per kg. or aminopyrine, 50 mg. per kg., was given subcutaneously. Controls were given intravenously a solution of sodium carbonate-bicarbonate buffer equivalent in ratio, molarity and volume to the duomycin pH 8.5. Additional controls received no drug. The range of weight for the rabbits was 1.5 to 3.0 kg.

Rats were made febrile by the subcutaneous injection of 10 cc. per kg. of a 15 per cent suspension of dried yeast.⁸ One group received orally 100 mg. per kg. of duomycin hydrochloride, a second group by the same route received 100 mg. per kg. of aminopyrine and the controls received a solution of hydrochloric acid equivalent in normality and total volume to the duomycin. The weights of the rats ranged from 200 to 300 grams.

apparently depends upon the fact that the pH and volume of the secretion prevent a prolonged contact.

Levels in Blood and Cerebrospinal Fluid. This study was confined to a convincing demonstration that duomycin passed the blood-brain barrier into the cerebrospinal fluid in therapeutically effective doses. No quantitative relationships between blood levels and cerebrospinal fluid levels were attempted. Dogs 468 and 200 were given 20 mg. per kg. intravenously at 8:40 A.M., 10:40 and 12:40 P.M. At 2:20 P.M., the arterial blood contained 40 gammas per cc. of serum, and at 2.30, the cerebrospinal fluid contained 0.8 gamma per cc. The results of tests on three other animals, dogs 273, 135 and 268, on the same schedule gave similar values. The cerebrospinal fluid from untreated dogs shows a titer less than 0.05 gamma per cc. Three undosed dogs were used.

Effect on Central Nervous System. In no animal was any effect on the central nervous system observed which could be attributed to duomycin as distinct from the acid or alkali administered with it. Doses of 50 mg. per kg. given intravenously at a rate of 10 mg. per kg. per minute to guinea pigs, rats, rabbits and dogs and at a rate of 150 mg. per kg. per minute to mice were tolerated with essentially no symptoms. Latent anticonvulsant or convulsant action was tested by determining the effect of 50 mg. per kg. intravenously in mice on the mortality produced by 100 mg. per kg. subcutaneously of metrazol. Of the 64 mice on metrazol alone, all convulsed and 44 per cent died. Of the 60 mice on metrazol, followed in 3 minutes by duomycin, all convulsed and 58 per cent died. This difference is not significant.

Summary and Conclusions

1. Duomycin, also known as aureomycin and as A-377, has a low toxicity and almost no side reactions.

2. Orally, mice tolerated 1500 mg. per kg. and rats 3000 mg. per kg.

3. The intravenous L.D.₅₀ for duomycin hydrochloride in mice was 134 mg. per kg., and for rats, 118. The alkaline form, pH 8.5, showed a toxicity of the same order.

4. Dogs, cats, rabbits, guinea pigs and mice tolerated without symptoms intravenous doses of 50 mg. per kg. pH 8.5 given at a rate of 10-20 mg. per kg. per minute. There was no evidence of methemoglobin formation.

5. Multiple intravenous doses of 20 mg. per kg. given to dogs two times per day for 6 days produced no unfavorable results except irritation of perivascular tissues at the site of injection.

6. Subcutaneous and intramuscular injections were irritating but 0.5 per cent solutions in 0.9 per cent saline produced only mild irritation in the conjunctival sacs of rabbits.

7. Mice, rats and dogs given 100 to 200 mg. per kg. per day orally for twelve weeks showed no evidence of chronic toxicity. The criteria were

growth, general appearance, hematology, clotting time, liver function, blood sugar, blood pressure and kidney function.

8. Duomycin pH 8.5 given to dogs intravenously at a rate of 10 mg. per kg. per minute produced essentially no changes in blood pressure or respiration. The hydrochloride, pH 2.5, was tolerated almost as well. There was one important difference. Doses of 30 to 40 mg. per kg. of the hydrochloride or an equivalent quantity of hydrochloric acid produced hemoglobinuria. This reaction was never observed with duomycin pH 8.5 even at 100 mg. per kg. Cats tolerated intravenous doses of the same order as those given to dogs.

9. Duomycin did not modify the vasomotor action of epinephrine, acetylcholine, or histamine; or the effect of vagal stimulation upon the heart.

10. Doses of 5 to 50 mg. per kg. of duomycin pH 8.5, given intravenously at the rate of 5 to 10 mg. per kg. per minute, did not modify appreciably the electrocardiograms of dogs.

11. Duomycin is a mild diuretic, about one third as active as caffeine. It does not produce albuminuria.

12. It has no effect upon blood sugar, isolated intestine or uterus and it does not potentiate or inhibit histamine.

13. It is not an antipyretic in rabbits or rats.

14. After oral doses, it appears in the urine in one hour and its excretion continues actively for 6 to 12 hours.

15. Therapeutically effective concentrations exist in the cerebrospinal fluid within 6 hours after an intravenous dose.

Bibliography

1. DUGGAR, B. M., & C. W. HESSELTINE. 1948. Characteristics of *Streptomyces aureofaciens*. Bull. Torrey Botanical Club. (In press.)
2. WENDEL, WILLIAM B. 1938. Methemoglobin determination. J. Lab. & Clin. Med. 24: 96.
3. SOMOGYI, MICHAEL. 1937. A reagent for the copper-iodometric determination of very small amounts of sugar. J. Biol. Chem. 117: 771.
4. LIPSCHITZ, W. L., Z. HADIDIAN, & A. KERPESAR. 1943. Bioassay of Diuretics. J. Pharmacol. & Exp. Therap. 79: 97.
5. SMITH, P. K., & W. E. HAMBOURGER. 1935. The ratio of the toxicity of acetanilide to its antipyretic activity in rats. J. Pharmacol. & Exp. Therap. 54: 346.
6. DORNBUCH, A. C., & E. J. FELCAK. 1948. The determination of Aureomycin in serum and other body fluids. Ann. N. Y. Acad. Sci. 51(2): 218-220.
7. KERSTEN, H., W. G. BROSENE, JR., F. ABLONDI, & Y. SUBBAROW. 1947. A new method for the indirect measurement of blood pressure in the rat. J. Lab. & Clin. Med. 32: 1090.
8. ROBBINS, BENJAMIN H. 1944. Dinitrophenol cataract: Production in an experimental animal. J. Pharmacol. & Exp. Therap. 80: 264.
9. LITCHFIELD, J. T., JR., & J. W. FERTIG. 1941. On a graphic solution of the dosage-effect curve. Bull. Johns Hopkins Hosp. 69: 276.
10. WILCOXON, F., & S. E. A. MCCALLAN. 1939. Theoretical principles underlying laboratory toxicity tests of fungicides. Contrib. Boyce Thompson Inst. 10: 329.

BACTERIOLOGICAL STUDIES OF AUREOMYCIN

By CLIFFORD W. PRICE, WILLIAM A. RANDALL, AND HENRY WELCH

*Division of Penicillin Control and Immunology,
Food and Drug Administration, Washington, D. C.*

PRELIMINARY studies of aureomycin have indicated that it is successful in the treatment of certain Rickettsial and viral diseases and, in addition, is effective against both the gram-positive and gram-negative types of organisms. In the early work on this drug, difficulty was encountered in the development of assay methods, particularly for the assay of serum concentrations in blood serum. It seemed important, therefore, to attempt to develop both cup-plate and serial dilution assays that could be depended on to evaluate aureomycin from the standpoint of both potency and its concentration in the blood following oral or parenteral use. Accordingly, sensitivities of a variety of organisms to the activity of aureomycin have been determined to establish the range of the bacterial spectrum for this drug and to identify organisms that would be useful in assay procedures.

The aureomycin* used in these investigations was relatively pure and was supplied as a sterile, amorphous hydrochloride. It was freely soluble in distilled water at a concentration of 2 per cent and produced a golden yellow solution having a pH of 4.5. All bacterial sensitivities were conducted in penicillin assay broth¹ utilizing the double serial dilution method, and a 1 per cent inoculum of a twenty-four hour broth culture of the desired organism. The tests were incubated at 37° C. for ninety-six hours, after which all negative tubes in the series were inoculated into fluid thioglycollate medium and incubated overnight to determine if any viable organisms remained. TABLE 1 presents these data, where it will be noted that gram-positive organisms are, in general, affected by much lower concentrations of aureomycin than are the gram-negative types. The most sensitive group studied appears to be the aerobic spore-bearing microorganisms. The variation between the bacteriostatic and the bactericidal endpoint was marked. When an organism is cultivated in the presence of aureomycin, inhibition of growth at the twenty-fourth hour may be brought about by very small concentrations, yet continued incubation results in growth in the tubes, and after incubation for ninety-six hours sub-culture in fluid thioglycollate broth reveals many surviving organisms in aureomycin concentrations hundreds of times greater than that inducing twenty-four hour inhibition of the same organism. Phenomena of this type could be due to the survival of resistant variants or

*Supplied through the courtesy of Lederle Laboratories Division, American Cyanamid Company, Pearl River, N. Y.

TABLE 1

	No of strains	Range of concentrations necessary to inhibit		Concentrations at which no viable cells remained
		24 hours	96 hours	96 hours
<i>Sal. typhosa</i>	6	78-1.56	25-50	25-50
<i>Sal. typhi murium</i>	1	1.56	50	50
<i>Sal. pullorum</i>	2	.39	12	25-50
<i>Sal. derby</i>	1	3.12	50	200
<i>Sal. manhattan</i>	1	3.12	50	200
<i>Sal. oregon</i>	1	3.12	50	200
			5	
<i>E. coli</i>	4	1.56-3.12		25-200
<i>E. intermedium</i>	1	6.25	50	50
<i>St. hemolyticus</i>	4	19-3.12	6-25	25-50
<i>St. fecalis</i>	4	50	50	50-200
<i>S. aureus</i>	4	09-25	3-50	50
<i>S. citreus</i>	1	.09	12	25
<i>Kl. pneumoniae</i>	6	.78-12.5	6-100	50-200
<i>Pr. vulgaris</i>	1	3.12	50	50
<i>Ps. aeruginosa</i>	2	50		
<i>Sar. lutea</i>	2	.09-19	3-50	50
<i>Sar. marcescens</i>	3	6-50	100-200	100-200
<i>Alkaligenes, sp.</i>	1	.19	6	25
<i>B. subtilis</i>	4	02-1.56	.78-12.5	1.5-50
<i>B. megatherium</i>	1	.02	3.12	3.12
<i>Bacillus, sp.</i>	3	04-1.56	.78-6.25	1.5-25

to destruction of the antibacterial agent during the incubation period, or both.

Repeated transfers of representative organism were made in penicillin broth in order to determine whether a rapid increase in resistance to aureomycin could be induced. Increases of this type have been reported in the case of other antibiotics. Finland² reported that an increase in resistance of an organism *in vivo* during the administration of streptomycin often occurred within forty-eight hours, the increase in resistance being as much as 4,000-fold. One of the organisms employed in this study, *Salmonella typhosa*, showed an increase of only 4-fold in resistance to aureomycin after fourteen transfers, yet in an identical number of transfers through a medium containing streptomycin an increase of 226,000-fold was demonstrated.³ Penicillin assay broth was employed as the medium. Serial dilutions by halves were made of the aureomycin and a 1 per cent suspension of the organisms added. The first tube showing growth was used as the culture for making up the suspension for each succeeding transfer. The results are given in TABLE 2, where it can be seen that even though some resistance does develop, with the exception of *Proteus vulgaris*, this increase in resistance could not possibly explain the variation between the bacteriostatic and bactericidal endpoints.

The stability of aureomycin in the presence of various culture media

TABLE 2

Microorganism	Sensitivity		
	Beginning $\mu\text{g/ml}$	After 14 transfers $\mu\text{g/ml}$	Increase
<i>B. subtilis</i>	0.04	2.5	62
<i>Sal. typhosa</i>	1.2	5.0	4.1
<i>E. coli</i>	0.6	10	16.6
<i>Kl. pneumoniae</i>	2.5	80	32
<i>Pr. vulgaris</i>	1.2	640	533
<i>Sar. lutea</i>	.15	.3	2
<i>S. aureus</i>	.08	.08	0
<i>St. hemolyticus</i>	.15	10	66.6

and culture media ingredients was then studied to determine if certain compounds were responsible for a deterioration of the antibiotic. The concentration of aureomycin employed was 20 $\mu\text{g/ml}$ and the deterioration, after incubation at 37° C., was measured by means of a *Bacillus subtilis* cup-plate technique to be described. Results are shown in TABLE 3. It thus appears that at 37° C. aureomycin is unstable even in water,

TABLE 3

Material tested	Concentration	% Loss in potency	
	%	5 hours	24 hours
Penassay broth	—	43	96
Beef extract	0.15	35	90
Peptone	.5	24	76
Phos. buffer pH 7.0	.5	56	89
Fl. thioglycollate	whole medium	32	60
Casitone	.15	25	80
Yeast extract	.5	40	80
Dextrose	.5	8	46
Sodium chloride	.25	25	82
L-cystine	.1	20	51
Na thioglycollate	.05	20	50
Trypticase	1.5	20	55
Thiol medium (Difco)	3.0	25	98
N-Z case peptone	1.5	37	90
Water, distilled	—	15	46

and this instability is markedly increased in the presence of a variety of compounds and media. Penicillin assay broth, Thiol medium, experimental (Difco), various peptones, beef extract, and phosphate buffer at pH 7.0 resulted in almost complete destruction after twenty-four hours' incubation. Fluid thioglycollate medium, on the other hand, resulted in a destruction only slightly greater than that obtained with distilled water. Certain constituents of the fluid thioglycollate medium, when used in the concentration at which they appear in the finished medium, gave very marked loss of potency. Combinations of the various ingredi-

ents to determine which were responsible for the apparent stabilization afforded by the complete medium are under investigation.

Various organisms were used in an attempt to develop a cup-plate assay method for aureomycin. One of these, *B. subtilis* 219, which is the organism used routinely in the assay of streptomycin,¹ was found to be the most satisfactory. The method is described below.

Prepare agar plates from penicillin assay agar¹ exactly as for the assay of penicillin but use *B. subtilis* PCI-219 for inoculating the seed layer. Place six standard penicillin assay cylinders on the seeded surface and fill the cups alternately with a dilution of unknown and with a reference standard made to contain 1.0 μg /ml. All unknowns and standard are dissolved or diluted with citrate buffer pH 6.3. Incubate the plates at 37° C. for eighteen hours and measure the zones of inhibition. The potency of the unknown solution is calculated using a standard curve. A typical assay for developing a standard curve for aureomycin is given in TABLE 4. All final dilutions are made in citrate buffer pH 6.3, since the use of a phosphate buffer not only rapidly inactivates aureomycin but results in "fuzzy" zones which are difficult to read.

TABLE 4

Concentration of aureomycin μg /ml.	Zone of inhibition mm.
0.1	14.5
0.4	19.3
0.6	20.7
0.8	21.7
1.0	22.7
1.2	23.3
1.5	24.1

The cup-plate assay method described was not applicable to the determination of aureomycin in blood serum because of poor zone formation and inconsistency in zone size. Serial dilution methods were then investigated using *B. subtilis* PCI-219 in a variety of media. The organism was found to be more sensitive to the action of aureomycin in fluid thioglycollate medium. This medium was also the one which resulted in the least destruction when aureomycin was incubated in its presence. The following test was devised.

Place 0.5 ml amounts of fluid thioglycollate medium in Wassermann tubes and serially dilute, by halves, the serum under test for the desired number of tubes. Place 0.5 ml of the serum in an empty Wassermann tube and place it first in the series. Prepare a standard for comparison by dissolving aureomycin in normal serum to give a concentration of 10 μg /ml and prepare a duplicate fifteen tube series exactly as above. A one μg . standard can be used if it is made up fresh, kept under refrigeration, and used in the test within a few hours. Maintain the *B. subtilis* in penassay broth (Difco). Add 0.5 ml of a 1:100,000 twenty-four hour broth culture of the *B. subtilis* to each tube and incubate overnight at 25° C. The last tube showing no growth is considered the endpoint, and the concentration in the unknown is determined by comparing with the standard aureomycin series. Approximately 0.01 μg /ml of aureomycin is demonstrable with this procedure. A representative determination is given in TABLE 5.

TABLE 5

Tube No.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	Concentration
Standard	-	-	-	-	-	-	-	-	-	-	-	+	+	+	+	
Standard	-	-	-	-	-	-	-	-	-	-	-	+	+	+	+	
Serum	-	-	-	-	-	-	-	-	-	-	-	+	+	+	+	10.0 $\mu\text{g}/\text{ml}$
Serum	-	-	+	+	+	+	+	-	-	-	-	+	+	+	+	0.02 $\mu\text{g}/\text{ml}$
Serum	-	-	-	-	-	-	-	-	-	-	+	+	+	+	+	5.0 $\mu\text{g}/\text{ml}$
(-) No Growth											(+) Growth					

Animal protection tests were conducted using six microorganisms. All injections were made intraperitoneally, utilizing 0.5 ml of the culture, followed immediately by a single appropriate dose of aureomycin in 0.5 ml of distilled water. The results are given in TABLE 6.

TABLE 6

Microorganism	Aureomycin μg .	No. of mice	Per cent deaths
<i>S. typhosa</i>	500	20	0
	250	20	10
	125	20	25
	50	40	35
	40	40	85
	25	40	100
	10	60	100
	0	60	99
<i>S. typhi murium</i>	500	20	90
	250	20	90
	125	20	95
	0	20	95
<i>E. coli</i>	500	20	0
	250	20	10
	125	20	10
	10	20	10
	0	20	70
<i>Kl. pneumoniae</i>	50	20	100
	25	40	100
	0	20	100
<i>Pr. vulgaris</i>	50	20	15
	25	20	20
	10	20	70
	0	20	100
<i>St. hemolyticus</i>	50	20	35
	25	20	50
	10	20	60
	0	20	100

Aureomycin behaved somewhat differently than penicillin or streptomycin when the latter drugs were employed in an identical experiment. When penicillin or streptomycin was used, the animals surviving were normal by the forty-eighth hour, while many of the animals treated with

aureomycin died four or more days after infection. When the effectiveness of aureomycin against experimental *S. typhosa* infections is compared with the effectiveness of streptomycin against the same infection on a weight basis, it is about 1/100th as active.

An LD₅₀ was determined, using white mice. All injections were by way of one of the tail veins, using a five second injection period. The results appear in TABLE 7.

TABLE 7

Dose		Spasms	Deaths
Per 20 gm.	Per kilo		
1.50 mg	75.0 mg	4 10	0 10
1.75	87.5	8 10	4 10
2.00	100.0	10 10	5 10
2.25	112.5	10 10	7 10
2.5	125.0	10 10	7 10
2.75	137.5	10 10	10 10
3.0	150.0	10 10	10/10

$$LD_{50} = 10 \pm 3 \text{ mg kg} \pm 0.3 \pm \text{mg}$$

The sample employed was found to be free from pyrogenic substances. When 20 mg. kg. of rabbit was injected the average temperature rise observed was 0.3° C. This dose, far in excess of the 1.0–2.0 mg./kg. of penicillin and 10.0 mg. kg. of streptomycin given in routine pyrogen testing, was well tolerated.

The effect of aureomycin on the blood pressure was studied in anesthetized cats. Phenobarbital was used as an anesthetic and the arterial blood pressure was determined by directly cannulating the carotid artery. When 10 mg. kg. was administered, a slight but transient fall in blood pressure occurred. At a dose of 20 mg. kg. the fall in blood pressure was more marked but a rapid recovery occurred with no evidence of respiratory embarrassment in any of the animals. When the dose was increased to 40 mg. kg., precipitous falls in blood pressure occurred, accompanied by cardiac arrest. The respiration became very shallow but continued for some time. In one animal, following a 40 mg. kg. dose which induced an 80 mm. drop in pressure, cardiac arrest occurred, but a return to normal was obtained following manipulation of the heart. A second dose of this size, however, resulted in a precipitous fall in the blood pressure of 100 mm., after which the animal could not be revived.

When aureomycin was combined with penicillin in the treatment of experimental *S. typhosa* infections in white mice, a synergistic effect was observed ranging from 5–40 per cent of that expected from the additive effects of these two drugs. When aureomycin was combined with streptomycin, however, in similar experiments, the results were not consistent.

Discussion

It is apparent from these studies that the bacterial spectrum of aureomycin is somewhat similar to that of streptomycin, inasmuch as it has an inhibitive action against both gram-positive and gram-negative

organisms. Its greatest activity appears to be against the gram-positive spore-bearing organisms. In general, it may be said that its activity is not as great as that of streptomycin, but its acute toxicity appears to be of approximately the same order of magnitude. On the basis of animal experimentation, the drug is well tolerated, although it cannot be utilized in concentrations on a weight basis as great as those that may be utilized with penicillin. This investigation indicates that aureomycin in solution is an unstable antibiotic and loses potency rapidly in broth, water and serum. Various media ingredients cause marked destruction. Such substances as peptone, beef extract, and even phosphate buffer at pH 6.3 materially reduce the activity of the drug in a relatively short period of time. It is apparent that the use of a thioglycollate medium tends to preserve the potency of aureomycin to a greater degree than the other media utilized. This is of importance in the bio-assay procedure for determining serum blood concentrations of this antibiotic. The assay procedure originally utilized in these studies was unsatisfactory, since the medium as a whole or the media ingredients caused rapid destruction of the drug during the testing procedure. This destruction has been considerably overcome by the assay procedure described using thioglycollate medium and incubation at 25° C. Under these conditions, it is possible to determine concentrations of aureomycin of approximately 0.01 µg. per ml. in serum. No outstanding usefulness for this antibiotic has been demonstrated by these animal investigations against the bacteria studied, although a definite synergistic effect has been demonstrated against certain strains of organisms with penicillin. The greatest promise for the drug thus lies in the treatment of certain viral and Rickettsial diseases.

Summary and Conclusions

1. The bacterial spectrum of aureomycin has been determined using a variety of gram-positive and gram-negative bacteria.
2. Both a serial dilution method of assay and a cup-plate assay have been developed.
3. In solution, aureomycin appears to be quite unstable and is affected by a variety of substances.
4. In the treatment of bacterial infections in mice, aureomycin is not as effective as either streptomycin or penicillin.
5. Aureomycin is a relatively non-toxic drug and is well tolerated by the animals used in these studies.

References

1. Federal Register, April 4, 1947, 12 F.R. 2215, 2217-2226 Title 21, part 141. Tests and Methods of Assay for Antibiotic Drugs Penicillin and Streptomycin.
2. FINLAND, M., R. MURRAY, H. W. HARRIS, L. KILHAM, & M. MEADS. 1946. Development of streptomycin resistance during treatment. J. A. M. A. **132**: 16-21.
3. PRICE, C. W., W. A. RANDALL, V. L. CHANDLER, & R. J. REEDY. 1947. Observations on the *in vivo* and *in vitro* development of bacterial resistance to streptomycin. J. Bact. **53**: 481-488.

THE DETERMINATION OF AUREOMYCIN IN SERUM AND OTHER BODY FLUIDS

By A. C. DORNBUSH AND E. J. PELCAK

*Lederle Laboratories Division, American Cyanamid Company
Pearl River, N. Y.*

THE potencies of antibiotic substances have been measured by several different methods. The most simple and perhaps the most widely used technique is that involving the serial dilution of the sample with agar¹ or with broth^{2, 3} and the noting of the last dilution at which complete inhibition of the test organism occurs. Turbidimetric assays⁴ are merely refinements of the dilution assays wherein the antibiotic substance is diluted in such a manner as to produce an end-point which will fall intermediate between complete inhibition and no inhibition. Agar diffusion methods⁵ similarly involve applying appropriate dilutions of the antibiotic in question to an inoculated agar surface and, after suitable incubation periods, measuring the zones of inhibition. With any assay method, it is important that a control solution be assayed in parallel with the "unknown" samples in order to have a point for comparison.

When the above methods were used in an attempt to assay aureomycin solutions, it was observed that the agar diffusion techniques as we applied them failed to produce sharp zones of inhibition. Also, the slope of the inhibition response curve was so low as to indicate a rather low order of precision. In making a survey of the various organisms readily available to us in the laboratory, we noted differences in the sensitivity of these organisms to aureomycin when tested by dilution methods. Since in selecting a test organism we must consider factors other than sensitivity alone, we eliminated those organisms which might be regarded as pathogenic. We did this not only for reasons of safety of laboratory personnel, but since we were primarily interested in assaying blood specimens we wished to eliminate the possibility of interference from specific bacteriostatic substances present in blood. In addition to the authentically identified cultures from our collection, we tested a number of laboratory isolates. One of these, *Bacillus* No. 5, seemed to meet our specification. It is a gram-positive spore-forming bacillus and appears to be a strain of *Bacillus cereus* variety *mycoides*.⁷

The method finally adopted for the assay of serum samples is similar to the methods originally proposed by the Food and Drug Administration for the assay of penicillin⁸ and streptomycin⁹ in blood. One-half ml. of a 0.2 $\mu\text{g.}/\text{ml.}$ solution of an aureomycin reference solution is placed in a $\frac{1}{2} \times 4\frac{1}{2}$ inch Wassermann tube. A second $\frac{1}{2}$ ml. portion of the same solution is diluted serially two-fold through an additional 5 tubes with sterile nutrient broth.¹⁰ Samples submitted for assay are similarly diluted

for as many tubes as may be required to give an end-point. If the approximate potency of the sample is known, it may be diluted with broth to contain approximately $0.2 \mu\text{g. ml.}$ after which it is serially diluted. When all of the samples and the standard have been placed in tubes, $1\frac{1}{2}$ ml. of a 1:100 broth dilution of an overnight broth culture of the test organism is added with an automatic pipette and the tubes are incubated at 37° in a water bath. After 4 hours, the tubes are examined visually and the growth is recorded as either negative or positive. The end-point is usually quite easily distinguishable and even faint growth is recorded as positive. Quite consistently, we have had an end-point in the third tube of the standard series; thus, the smallest amount of aureomycin we are able to detect is $0.05 \mu\text{g. ml.}$ of sample. Actually, the aureomycin concentration in the final clear tube is one-quarter as much or $0.0125 \mu\text{g. ml.}$, since $\frac{1}{2}$ ml. amounts have been made up to a volume of 2 ml. On several occasions, we have read the tubes at the 4-hour interval and then have incubated the tubes an additional 16 hours. Generally, the sensitivity of the assay is lowered to one-half or to one-quarter of that found at the 4-hour reading, but the actual potencies found in the samples do not change since the shift is parallel in the sample tubes and in the standard tubes. In the case of non-sterile urine specimens, when the assay tubes are incubated for a total of 20 hours, growth of the contaminating organisms may partially obscure the end-point. Even with grossly contaminated samples, however, we have usually been able to read end-points at 20 hours by observing the point at which the characteristic pellicle of *Bacillus* No. 5 occurs in each series. Because of the massive inoculum used initially, we have not had any interference from contaminated samples when they were read at the 4-hour interval. A total of 38 normal human urine specimens and 23 normal human serum specimens were examined with no indication of interference from inhibitory materials. Similarly, when we have fortified serum with aureomycin we have had no indication of interference from growth stimulation under the conditions of the assay.

Some of the early assay of blood and urine samples were carried out with *Bacillus subtilis* A.T.C.C. 6633. In these assays, the technique was identical with that used at present, but the tests were not as sensitive nor was it possible to read the tubes at the 4 hour time interval.

In addition to the dilution assay outlined above, we have utilized a turbidimetric assay for the determination of aureomycin in urine. As in the case of the dilution assay, we have modified existing penicillin techniques.⁵ The urine samples are diluted with 0.1 M monobasic potassium phosphate to contain approximately $0.1 \mu\text{g.}$ of aureomycin per milliliter. Aliquots of 0.2, 0.4, 0.6, 0.8 and 1.0 ml. of this solution are added to unplugged 18 x 150 mm. test tubes and the volume of each is adjusted to 1.0 ml. with the same 0.1 M phosphate using a Cannon automatic dispenser. An aureomycin reference solution is included in each assay and it is set up at ten levels in a range from $0.01 \mu\text{g.}$ to $0.1 \mu\text{g.}$ per tube. The test

organism, *Staphylococcus aureus* F.D.A. 209P, is carried out on nutrient agar slants. On the day of the assay, growth from a fresh slant is suspended in 150 ml. of nutrient broth,¹⁰ and the inoculated flask is maintained at 37° for exactly 1½ hours. A bottle containing the calculated amount of broth for the entire assay is inoculated with this culture at the rate of 35 ml. per liter. Nine-ml. amounts of the inoculated medium are dispensed in each assay tube with an automatic pipetting machine. After incubating exactly 3 hours at 37°, ½ ml. of formalin 1 : 3 is added to each tube and the turbidities are read on a photoelectric colorimeter.

Twelve normal human urine samples were pooled and aliquots of the pooled material were fortified with different amounts of aureomycin. When tested by both of our assay methods, the recoveries were satisfactory. Similarly, we have had good agreement between the parallel assays performed on various animal urine samples submitted to us by the pharmacology group.

Using the turbidimetric assay we have made some aureomycin stability determinations. The antibiotic is most stable near pH 2. An unbuffered aureomycin hydrochloride solution at pH 2.9 has been maintained at cold room temperature $\pm 4^{\circ}$ C. for 23 days with no measurable loss of activity. Temperature is an important consideration. Whereas aureomycin solutions from pH 1 through pH 10 have been held at $\pm 4^{\circ}$ for approximately 18 hours with no loss in activity, at pH 7.0, more than half of the activity is lost after overnight incubation at 37°.

Bibliography

1. WAKMAN, S. A., & H. C. RILEY. 1945. Ind. & Eng. Chem., Anal. Ed. **17**: 556.
2. FLOREY, H. W., & M. A. JENNINGS. 1942. Brit. J. Exp. Path. **23**: 120.
3. RAMMELKAMP, C. H. 1942. Proc. Soc. Exp. Biol. & Med. **51**: 95.
4. FOSTER, J. W. 1942. J. Biol. Chem. **144**: 285.
5. McMAHON, J. R. 1944. J. Biol. Chem. **153**: 249.
6. ABRAHAM, E. P., E. CHAIN, C. M. FLETCHER, H. W. FLOREY, A. D. GARDNER, N. G. HEATLEY, & M. A. JENNINGS. 1941. Lancet **241**: 177.
7. SMITH, N. R., R. E. GORDON, & F. E. CLARK. 1946. U. S. D. A. Misc. Pub. **559**.
8. RANDALL, W. A., C. W. PRICE, & H. WEICH. 1945. Science **101**: 365.
9. PRICE, C. W., J. K. NIELSON, & H. WILCH. 1946. Science **103**: 56.
10. SCHMIDT, W. H., & A. J. MOYER. 1944. J. Bact. **47**: 199.

IN VITRO STUDIES WITH AUREOMYCIN*

By CAROLINE A. CHANDLER AND ELEANOR A. BLISS

*Department of Preventive Medicine and Pediatrics,
Johns Hopkins University School of Medicine,
Baltimore, Maryland*

INVESTIGATIONS on aureomycin have been in progress in this laboratory since November 1947. The *in vitro* activity of this new antibiotic against various bacteria has been compared with the activity of penicillin, polymyxin, and streptomycin, and the results are being reported elsewhere.

To recapitulate these results briefly, it was found that, in the test tube, aureomycin is less effective than polymyxin and penicillin against the gram negative bacilli and the gram positive cocci, respectively. The only exception noted was in the case of *Streptococcus faecalis*, six strains of which were found to be more susceptible to aureomycin than penicillin. For gram negative bacilli including *E. typhosa*, *E. coli*, *K. pneumoniae*, *A. aerogenes*, and *H. influenzae* β , aureomycin was found to be inhibitory at levels of from 2.5 to 5.0 micrograms per cc. of medium. Both *B. proteus* and *B. pyocyaneus* were found to be relatively resistant to aureomycin as neither one was inhibited by concentrations of 20 micrograms per cc. All of the gram positive cocci tested, including several strains each of staphylococci, pneumococci and hemolytic streptococci, were inhibited by less than one microgram of aureomycin per cc.

In a few instances, comparison with streptomycin was also made. Against *E. coli* and *K. pneumoniae*, aureomycin was not as effective as streptomycin. One of the test strains of staphylococci was, however, more susceptible to the new antibiotic than streptomycin. Both agents were equally effective against one strain of *H. influenzae* β .

In the course of the above studies on aureomycin, evidence began to appear suggesting that this antibiotic is bacteriostatic rather than bactericidal in action and that it deteriorates rather rapidly in the usual test media. The following investigations were undertaken in an attempt to throw more light on these phenomena.

Methods

1. Growth curves were obtained by making bacterial counts at various intervals after inoculation. The test organism used in this experiment was a beta hemolytic streptococcus, strain C203. Penicillin was chosen as the agent of comparison. Ten cc. volumes of both agents were inoculated

* These investigations were supported by grants from Abbott Laboratories; Eli Lilly and Company; Lederle Laboratories Division of the American Cyanamid Company; Parke, Davis and Company; and the Upjohn Company.

with 0.1 cc. of 1:100 dilution of the culture and count plates were made at 1, 3, 5, 24, and 48 hours.

2. The susceptibility of the test strains to aureomycin was determined in the following manner. Serial two-fold dilutions of the antibiotic in Difco heart infusion broth, with 0.075 per cent dextrose added, were inoculated with equal volumes of 18-hour broth cultures so diluted as to result in an inoculum of approximately 200,000 organisms per cc. The tests were read at the end of 6, 12, 24, 48, 72, and 96 hours incubation at 37° C.

3. Drug deterioration was investigated in another set of experiments in which *K. pneumoniae* was the test organism. Three sets of tubes each containing two-fold dilutions of aureomycin from 0.625 to 20 micrograms per cc. were inoculated in the usual manner with the test strain. After 24 hours incubation growth was found to be inhibited by 2.5 micrograms per cc. of drug in all three sets of tubes. The first set (a) was allowed to incubate without further addition of drug through 72 hours. Fresh drug, to give an additional 2.5 cc. micrograms per cc. per tube, was added to the second set (b) at 24 hours, and to the third set (c) at 24 and 48 hours.

4. The activity of aureomycin in the presence of defibrinated, whole blood, fresh serum and heat-inactivated serum was determined in the following manner. Twenty-five hundredths cc. of blood or serum was introduced into each of 6 sterile pyrex tubes. Aureomycin, diluted in heart infusion broth to give the final concentrations desired, was added in 0.1 cc. amounts. Finally, 0.05 cc. of each ten-fold serial dilution (10^{-1} to 10^{-6}) of an 18-hour broth culture of the test organism was also added to each tube. The tubes were sealed, placed in a rotating box, and incubated at 37° C. for 24 hours, after which the tubes were opened and the contents cultured on blood agar plates. To determine the number of organisms added to each tube of blood or serum, count plates were made of the 10^{-4} and 10^{-5} dilutions. Control experiments were set up in the same fashion except that broth was used as the substrate rather than blood or serum.

Results

1. The growth curve, shown in FIGURE 1, indicates that although aureomycin brought about a progressive decline in the bacterial population for the first 24 hours of exposure, the organisms eventually recovered and attained growth equal to that in the control culture. Penicillin, on the other hand, caused an immediate decline in the bacterial count which terminated in the sterilization of the culture.

2. Whether or not the eventual outgrowth of organisms in the presence of aureomycin is the result of the survival of resistant organisms or is due to a deterioration of the drug is not known at present. That drug deterioration is an important and perhaps the most significant factor is illustrated in FIGURE 2. When aureomycin was titrated, in the manner already described, against *K. pneumoniae* as the test organism, the visible end-point shifted with progressive incubation so that at 6 hours the end-

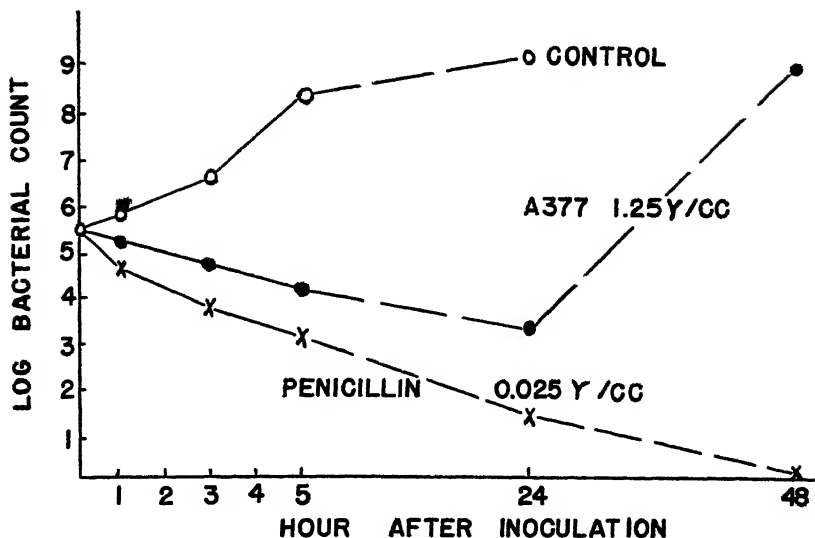


FIGURE 1 Effect of aureomycin and penicillin on the multiplication of 203

point was 0.3 microgram per cc.; at 12 hours it was 1.25 micrograms per cc.; at 24 hours it was 5.0 micrograms per cc.; at 48 hours, 10 micrograms per cc.; at 60 hours, 20 micrograms per cc.; and at 72 hours it was greater than 20 micrograms per cc. Since repeated tests with this organism always titrated to essentially the same end-point at any given time, a six-hour test was developed for assaying the content of aureomycin in the body fluids of patients to whom the drug was being administered therapeutically. In each instance *K. pneumoniae* was employed as the test organism and a variety of specimens of blood sera and urines were assayed, the results of which will be reported in a later paper in this series.

Confirmatory evidence of drug deterioration was obtained in another set of experiments. The manner in which these experiments were performed has been described (see *Methods 3*), and the results are shown in FIGURE 3. In three identical sets of tubes containing two-fold dilutions of aureomycin from 0.625 through 20 micrograms per cc.—all inoculated at the same time—growth was found to be inhibited by 2.5 micrograms per cc. in all three sets of tubes at the end of 24 hours incubation. With fresh drug added to the second set (b) at 24 hours, growth was held down to 2.5 micrograms through 48 hours and then titrated to more than 20 micrograms at 72 hours. In set (c), with drug replenished at both 24 and 48 hours, the original titer of 2.5 micrograms per cc. was maintained throughout the 72-hour observation period. These results are in marked contrast with set (a) where, with no replacement of drug, growth titrated to 20 micrograms at 48 hours and to more than 20 micrograms at 72 hours.

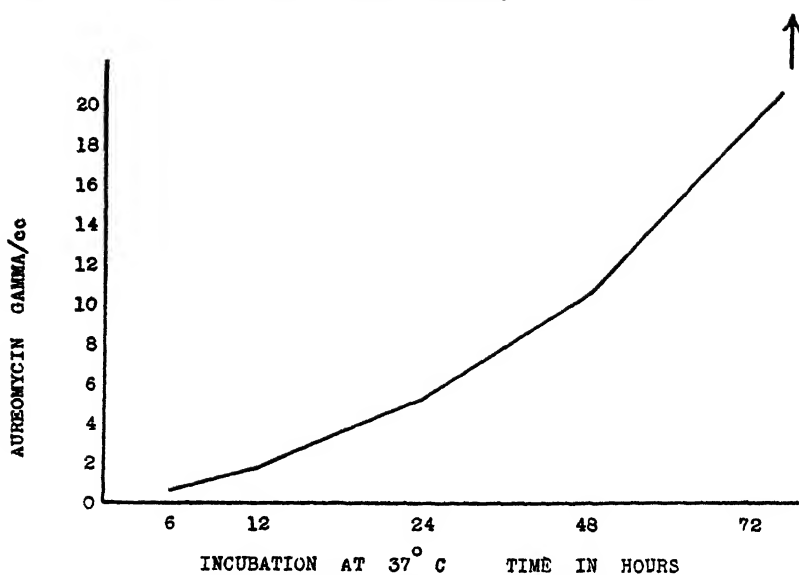


FIGURE 2. Shift in titer of aureomycin with progressive incubation. Test organism *K. pneumoniae*.

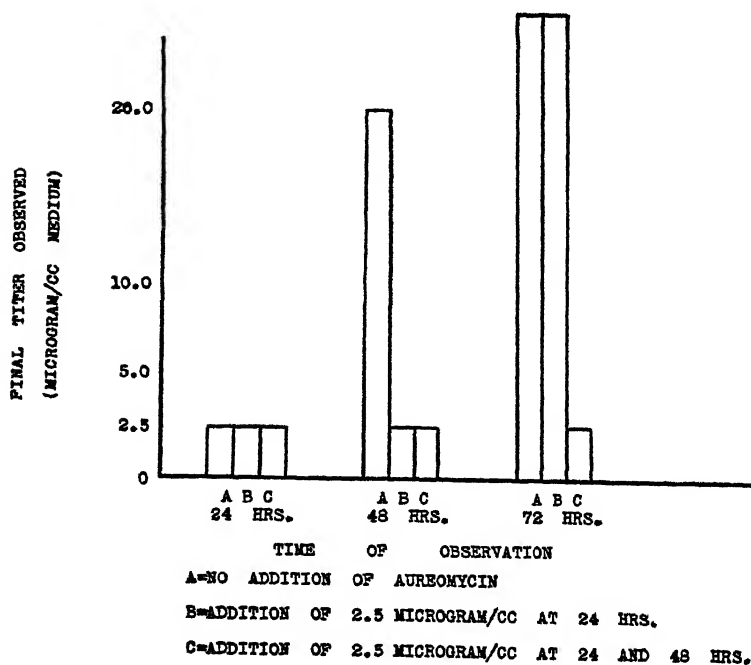


FIGURE 3. Effect of adding increments of fresh aureomycin.

That pH and temperature of the environment are significant factors in determining the rate of deterioration was shown by observing the end-points obtained with *E. coli* in solutions of aureomycin at different pH's held at -10° C. and at 37° C. for 23 hours previous to inoculation. The results are shown in TABLE 1. It can be seen that the deleterious effect of heat was augmented by an increase of pH.

TABLE 1

EFFECT OF THE TEMPERATURE AND pH OF STORAGE ON THE DETERIORATION OF AUREOMYCIN. TEST ORGANISM: *E. coli*; 24 HR. READINGS

Storage temperature	Storage pH			
	5.1	5.8	7.1	8.2
	Minimal inhibitory concentration— γ /cc			
-10° C.	3.12	3.12	6.25	6.25
37° C.	6.25	6.25	50.0	50.0

3. Studies on the activity of aureomycin in the presence of whole blood and serum are summarized in TABLES 2 and 3. When a *Staphylococcus*

TABLE 2

BACTERICIDAL STUDIES

Organism	Substrate	Minimum number of organisms required to initiate growth/cc. of blood or serum			
		Aureomycin gamma/cc.			
		0	5	10	20
<i>Staphylococcus aureus</i>	Whole blood	640	64,000	100,000	100,000
	Serum (fresh)	800		8,000	800,000
	Serum (inactivated)	800		80,000	800,000
	Broth (control)	1,600	1,600,000	1,600,000	1,600,000

TABLE 3

BACTERICIDAL STUDIES

Organism	Substrate	Minimum number organisms required to initiate growth/cc. of blood or serum		
		Aureomycin gamma/cc.		
		0	10	20
<i>K. pneumoniae</i>	Whole blood	3000	30,000,000	> 30,000,000
	Serum (fresh)	4000	40,000,000	> 63,000,000
	Serum (inactivated)	680	68,000	> 68,000,000
	Broth	800	80,000,000	> 80,000,000

aureus was used as the test strain, as shown in TABLE 2, there was no indication that the bacteriostasis achieved with aureomycin alone was augmented by the bactericidal activity of whole blood or of serum, either fresh or inactivated. On the contrary, both blood and serum had a somewhat neutralizing effect on the action of the drug. This is evidenced by

the fact that, in the presence of 20 micrograms of aureomycin per cc., only 100,000 organisms were required to initiate growth in whole blood and 800,000 in serum, whereas in broth alone, 1,600,000 organisms were needed to produce growth. In this type of bacteriostatic test, there was no significant difference between fresh and inactivated serum with this organism.

Essentially the same results were obtained when *K. pneumoniae* was employed as the test strain, as shown in TABLE 3. In the presence of 20 micrograms of aureomycin per cc., more than thirty million organisms were required to initiate growth in whole blood, more than sixty-eight million in serum, but in broth alone, the minimal inoculum was eighty million. Again, the somewhat antagonistic effect of blood and serum on the drug is indicated by the above results. It is interesting to note that in the presence of 10 micrograms of drug per cc., fresh serum was more bactericidal than inactivated serum against *K. pneumoniae*. This is shown by the fact that forty million organisms were required to initiate growth in fresh serum as compared with only sixty-eight thousand organisms in heat inactivated serum. The most likely explanation for this observation is that the bactericidal effect of fresh complement on a gram negative organism tends to offset the neutralizing effect of serum itself on the drug. That this explanation is probable is suggested by the finding that neither fresh nor inactivated serum had any bactericidal effect in the absence of drug.

Discussion

The performance of aureomycin in the test tube is not brilliant when compared with that of polymyxin or penicillin. When due consideration is given to the fact that this antibiotic deteriorates rapidly under experimental conditions, it is surprising that aureomycin appears to be much more promising clinically, even at minimal blood levels, than it does experimentally. One is led to suppose that the mode of action in the body may be different from that *in vitro*.

In spite of the antagonistic effect of whole blood and serum, it is significant that organisms are actually killed in these substrates when sufficient quantities of aureomycin are present.

Summary

1. The new antibiotic, aureomycin, is effective *in vitro* against both gram negative and gram positive bacteria. Higher concentrations of this drug are required to produce bacteriostasis than are necessary with either polymyxin or penicillin, being close to those required of streptomycin.
2. Aureomycin is bacteriostatic rather than bactericidal in its effect.
3. The drug deteriorates rapidly at room and incubator temperatures in neutral or alkaline solution, but bacteriostasis can be maintained if fresh drug is added every 24 hours. Since the rate of deterioration is

fairly constant for any given test organism, a rapid assay method can be used for determining the content of aureomycin in body fluids.

4. The activity of aureomycin is not enhanced by the presence of whole blood or serum. On the contrary, although actual killing takes place if sufficient drug is present, both blood and serum exert an antagonistic effect on the activity of this antibiotic

LABORATORY STUDIES WITH AUREOMYCIN*

By THOMAS FITE PAINE, JR.,† HARVEY SHIELDS COLLINS,
AND MAXWELL FINLAND

*Thorndike Memorial Laboratory, Second and Fourth Medical Services (Harvard)
Boston City Hospital, and the Department of Medicine, Harvard Medical School,
Boston, Massachusetts*

SOME of the results of the studies carried out by the workers of the Lederle Laboratories which were made available to us suggested that aureomycin, a new antibiotic, was relatively non-toxic and highly effective, particularly against certain viral and rickettsial infections in animals, and seemed worthy of clinical trials. The studies reported here were carried out as a part of the evaluation of the clinical effect of this agent in the treatment of certain human bacterial infections.

Sensitivity of Various Bacteria. Tests for sensitivity were performed on 183 strains of organisms recently isolated from patients. Tube-dilution and streak-plate methods were used for these tests. The strains of *Diplococcus pneumoniae*, *Neisseria gonorrhoeae*, *N. meningitidis* and *Streptococcus pyogenes* were all inhibited by 1 μ g. or less of aureomycin per ml. Staphylococci and most strains of gram negative bacilli, including species of *Salmonella*, were inhibited for the most part by concentrations of aureomycin in the range of 1 to 25 μ g. per ml. Strains of only 2 species, namely *Proteus vulgaris* and *Pseudomonas aeruginosa*, were moderately resistant to the action of aureomycin, requiring from 125 to 250 μ g. per ml. for complete inhibition.

Comparison with Other Antibiotics. No evidence of cross-resistance with other antibiotics was found. Penicillin- and streptomycin-sensitive and resistant strains of the same organism were found to be equally sensitive to the antibacterial activity of aureomycin. On a weight basis, aureomycin was found to be less effective than penicillin, but more effective than streptomycin in inhibiting coccal organisms and was about equally as effective as streptomycin against most gram negative bacilli.

Stability. Aureomycin hydrochloride as provided to us was found to maintain its potency on storage at room temperature in the dry powder in sealed ampules for at least 7 months. It also retained most of its activity when kept in solution in high concentrations in distilled water at 37°

* Aided by a grant from the United States Public Health Service.

Generous supplies of aureomycin hydrochloride and some of the data concerning its activity *in vitro* and in experimental animals were supplied by the Lederle Laboratories Division of the American Cyanamid Company, Pearl River, New York.

The laboratory studies were carried out with the technical assistance of Clare Wilcox, Janice M. Bryan and Paul F. Frank. The strains of pathogenic bacteria used in these studies were isolated and identified by Marion E. Lamb and A. Kathleen Daley in the Bacteriology Laboratory of the Mallory Institute of Pathology and by Helen Trousdale in the Urological Clinic, Boston City Hospital.

Dr. E. Buist Wells participated in some of the pharmacologic studies.

† Research Fellow, American College of Physicians.

C. or at 4° C. In low concentrations in broth, plasma or blood agar, however, a rapid decrease in potency occurred at 37° C., and a somewhat less rapid fall in potency took place at 4° C

Comparison of Tube-dilution and Streak-plate Methods for Determining Sensitivity of Bacteria

Effect of Size of Inoculation. Bacterial sensitivities to aureomycin could be performed satisfactorily using either a tube-dilution technique or by streaking cultures directly on the surfaces of a series of agar plates containing graded concentrations of the antibiotic. The inoculum size influenced the concentration of aureomycin required for complete inhibition when using the tube-dilution method, but did not alter the inhibition end-points when the surface streak method of determining sensitivities was used. Aureomycin did not, apparently, exert antibacterial activity against fully grown cultures of microorganisms. However, against vigorously multiplying organisms, it exerted considerable antibacterial activity.

Effect of pH. The pH of the media exerted considerable influence on the inhibition end-points observed in the tube-dilution sensitivity tests. Increasing the pH from 6.1 to 8.0 resulted in a 64-fold increase in the concentration of the antibiotic required to inhibit one of the organisms tested.

"Aureomycinase"? There was no evidence of the elaboration of any aureomycin-inhibiting substance (similar to penicillinase) by any aureomycin-resistant organisms, nor could any such material be extracted from the disrupted bacterial cells of a number of organisms.

Effect of Aerobiosis. The degree of aerobiosis may affect the inhibition end-points of aureomycin against certain organisms. Preliminary experiments suggested that a marked reduction in oxygen content in the environment of a culture resulted in some diminution in the antibacterial activity of aureomycin. The activity of aureomycin on cultures that were incubated in a candle jar, however, was not diminished.

Development of Resistance IN VITRO. The bacteria studied showed no marked tendency to become resistant to aureomycin on repeated exposure to the drug, though an occasional strain did increase in resistance after many subcultures in increasing concentrations of the antibiotic. No evidence could be found for the development of aureomycin "dependent" strains of bacteria. Resistant variants could not be isolated readily by exposure of large bacterial populations to high concentrations of the antibiotic.

Absorption and Excretion

Blood Levels. Many attempts were made to determine the blood concentrations of aureomycin in patients receiving this antibiotic by assaying the serum against selected test strains of bacteria. Despite doses up to 2.0 grams daily by mouth, it was not possible to inhibit a test strain of

Streptococcus which was sensitive to 0.5 $\mu\text{g.}/\text{ml.}$ in serum dilutions beyond 1 : 2 or occasionally 1 : 4. With *Bacillus* No. 5 (a *subtilis*-like organism inhibited by 0.2 $\mu\text{g.}$ of aureomycin per ml.), it was possible to obtain inhibition in serum dilutions up to 1 : 8 or 1 : 16.

Urinary Excretion. Following oral or intramuscular doses of aureomycin, the material appeared in the urine within the first hour and the maximum amount appeared during the following 4 to 8 hour period. Following single doses of 0.5–0.75 gm. by mouth, as much as 13.4 per cent of the dose could be recovered in the urine during the following 55 hours. The maximum concentration in the urine on oral doses of 1–2 grams daily ranged up to 256 $\mu\text{g.}$ per ml. An interesting feature was the prolonged excretion of the drug, following oral doses. Antibiotic activity was demonstrated in urine collected after three, and sometimes after four days following the last oral dose. In patients receiving repeated doses for severe infections, the total amount recovered from the urine was only 3–5 per cent of the amount administered, but technical difficulties may have accounted, in part, for the failure to recover larger proportions of the administered drug.

Summary

Aureomycin possesses antibacterial activity against many microorganisms, including coccal and bacillary forms. It is active against certain penicillin-resistant organisms as well as against streptomycin-resistant and dependent strains. It is stable in the dry powder at room temperature and in high concentration in solution in distilled water at 37° C. or 4° C. Loss of activity occurs rapidly when the antibiotic is incubated at 37° C. in low concentration in broth, plasma or blood agar, though the loss in activity in the same media is less rapid on refrigeration.

The antibiotic is most active in the acid pH range. The number of organisms present, as well as their growth phase, influences the inhibition end-points observed in the tube-dilution method of sensitivity determination.

Bacteria do not readily become resistant to aureomycin *in vitro*, though such an alteration was achieved in the case of certain microorganisms.

On oral doses of aureomycin of 1 or 2 grams daily in adults, concentrations up to 256 $\mu\text{g.}$ per ml. may be found in the urine. Excretion of the antibiotic in the urine continues for 3 days or longer after its administration.

CLINICAL STUDIES WITH AUREOMYCIN*

By HARVEY SHIELDS COLLINS, THOMAS FITE PAINE, JR.,†
AND MAXWELL FINLAND

*Thorndike Memorial Laboratory, Second and Fourth Medical Services (Harvard)
Boston City Hospital, and the Department of Medicine, Harvard Medical School,
Boston, Massachusetts*

SOME of the preliminary findings on the toxicity of aureomycin and its effect in experimental infections that were made available to us by the workers of the Lederle Laboratories suggested that this agent was relatively non-toxic, was effective by mouth, and gave promise of being highly active against certain viral and rickettsial infections. Studies *in vitro* by the same workers, which were confirmed and extended in this laboratory, also indicated that it was effective against a wide range of human pathogenic bacteria. This new antibiotic, therefore, seemed worthy of clinical trials in bacterial infections in order to determine its range of usefulness.

Certain laboratory studies which were carried out as part of the clinical trials are reported separately.^{1, 2} To date, about 100 patients with a variety of bacterial infections have been treated with aureomycin by mouth at the Boston City Hospital. The results obtained in these patients are summarized in this communication.

Infections Treated and General Results

The types of infection treated, the dosage used, and a rough estimation of clinical effect of aureomycin in these cases are listed in TABLE 1. It should be borne in mind that most of these patients were treated before there was much knowledge of the toxicity of the drug in humans and the dosage employed was gradually increased in successive patients until considerably greater amounts were used than any that had previously been given by mouth. These patients received no other chemotherapy or antibiotic during the aureomycin treatment, although many of them had been treated unsuccessfully with such agents before the aureomycin was started.

Good results were obtained in most of the coccal infections. The effects in most of the patients with typhoid fever, *Salmonella* infections and severe urinary tract infections were difficult to evaluate. In most of the patients in whom the results are listed as doubtful, definite clinical and bacteriological evidence of beneficial effects was noted during treatment with aureomycin, but the results were either transient or the role of the antibiotic in effecting the results could not be evaluated.

* Aided by a grant from the United States Public Health Service.

† Research Fellow, American College of Physicians.

TABLE 1

SUMMARY OF RESULTS OF AUREOMYCIN TREATMENT IN 102 PATIENTS

Diagnosis	Patients	Total therapy, by mouth		Results		
		Grams	Days	Good	Doubt- ful	Failed
Gonorrheal urethritis, male	66	1-3	1-2	49	11	6
Pneumococcal pneumonia	1	5-20	5-10	4	0	0
Meningococcemia, acute	1	4	3	1	0	0
Typhoid fever	5	3-39	13-22	1	2	2
Typhoid carrier	1	23	31	0	0	1
Acute enteritis, <i>S. newport</i>	2	5-7	3-8	1	0	1
<i>S. suispestifer</i> bacteremia (vertebral and psoas abscesses)	1	21	11	0	0	1
<i>E. coli</i> bacteremia	1	22.5	15	0	0	1
Cystitis and pyelonephritis	16†	3-28	3-28	6	15	2
Lymphogranuloma venereum	1	40	20	0	1	0
Measles	1	1	2	0	0	1
Relapsing fever (?)	1	9-11	5-6	0	0	1
Non-specific urethritis	2	4	7	2	0	0
Total	102			64	29	16

* One of these patients also received 20 mg. intramuscularly twice a day.

† Seven of these patients received 2 separate courses because of reinfection with new strains and are each listed twice under "Results."

Suitable cases of rickettsial infections or of viral diseases other than those listed were not available for treatment during the study. A brief résumé will be given of the results in certain groups of cases.

Gonococcal Urethritis. The cases of acute gonococcal urethritis were all in males, and were treated in the Out-Patient Clinic. The results with the various dosage schedules that were tried are shown in TABLE 2. Failures were most frequent among the patients who were treated for only one day and with a total dose of only 1 or 1.5 gm. of aureomycin. In them, urethral discharge was present and contained numerous well-formed leucocytes on the day following treatment. This was true even among those listed as "Cured" and even when cultures and smears of this discharge were negative for gonococcus. Complete subsidence of the discharge in these cases occurred in 2 or 3 days. The patients who received doses on 2 successive days showed a response more like that which usually follows an adequate dose of penicillin. In them, the discharge cleared rapidly and leucocytes, when present on the day after treatment ended, looked disrupted and stained poorly.

It is thus evident that aureomycin was effective against acute gonorrhea in the males. The over-all results in this group of cases, however, are distinctly inferior to those observed in this same clinic in patients treated with 300,000 units of penicillin given intramuscularly in a prolonging agent. The optimum dosage schedule of aureomycin, however, has not yet been defined, and the effects of intramuscular therapy were not explored.

TABLE 2
SUMMARY OF RESULTS IN CASES OF ACUTE GONOCOCCAL URETHRITIS IN MALES
'TREATED WITH AUREOMYCIN BY MOUTH'

(Boston City Hospital -to July 9 1948)

Number of patients	Aureomycin		Total gm.	"Cured"	Failed†	No follow-up
	Schedule 0.5 gm. doses					
9	10 a.m. & 10 p.m.		1.0	4	3 (1)	2
21	10 a.m.; 5 & 10 p.m.		1.5	17	3 (1)	1
5	10 a.m.; 2, 6 & 10 p.m.		2.0	4‡	0	1
21	10 a.m. & 10 p.m. (day 1); 10 a.m. & 10 p.m. (day 2)		2.0	14	3 (1)	4
8	10 a.m.; 5 & 10 p.m. (day 1). 10 a.m. & 10 p.m. (day 2)		2.5	8	0	0
1	10 a.m. & 10 p.m. (day 1); 10 a.m. (day 2)		1.5	1	0	0
1	10 a.m.; 2, 6 & 10 p.m. (day 1);** 10 a.m. & 10 p.m. (day 2)		3.0	1	0	0
66				49	9 (3)	8

†Bacteriological studies carried out by Mrs. Helen Trousdale; the clinical studies were carried out through the courtesy of Dr. George C. Prather, Chief of the Urological Service, with the assistance of Drs. Thomas F. Kaiser and Francis C. Regan.

‡In parentheses are shown the number of patients, included among the failures, who had delayed reappearance of symptoms and positive cultures after having completely cleared, they may represent reinfections.

** Discharge present and culture positive at 9 a.m. on day 2.

‡ One patient had acute orchitis which subsided promptly.

Pneumococcal Pneumonia. Four patients with pneumococcal pneumonia were treated with aureomycin by mouth. The relevant findings are summarized in TABLE 3. In 3 of the patients, including one with bacteremia, treatment was begun on or before the second day of illness. In each of these patients improvement, both subjective and objective, occurred rapidly and they were all afebrile within 18 to 36 hours. In patient B. C., who had a very severe infection with jaundice, treatment with aureomycin was started on the fourth day. Improvement in this patient was somewhat more gradual and the temperature dropped to normal within 48 hours.

Careful bacteriological studies in these 4 patients revealed a rapid decline in the number of pneumococci in the sputum following the first dose of aureomycin and no pneumococci could be recovered in any of them after 48 hours. The results are similar to those obtained with full doses of penicillin in comparable patients.

The various organisms isolated from the sputum in patient J. T. and the sensitivity of these organisms to aureomycin are listed in TABLE 4. Only alpha hemolytic streptococci could be isolated after the second day of treatment.

Meningococcemia. The patient with meningococcemia is of interest because treatment with aureomycin was started within 12 hours of the onset of an illness which was tentatively diagnosed as Rocky Mountain spotted fever. The patient had been in an endemic focus and had had tick

TABLE 3
RELEVANT FINDINGS IN 4 PATIENTS WITH PNEUMOCOCCAL PNEUMONIA TREATED WITH AUREOMYCIN BY MOUTH

Patient	Age	Type	Blood culture	Lobes*	Severity	Dosage		Day treatment begun	Hours of fever after first dose
						Schedule	Total gm.		
PG	19	2	0	LL	++	0.5 gm. + 0.25 Gm. Q 6 Hr.	5.	2	36
BC	65	7	0	LL	+++	0.5 gm. Q 6 Hr.	20.	1	48
SW	23	4	0	RL	++	0.5 gm. Q 8 Hr.	10.5	1	18
JT	16	44	+	RL	+++	0.5 gm. Q 6 Hr.	10.	2	18

* LL = left lower; RL = right lower.

TABLE 4

SENSITIVITY TO AUROMYCIN OF STRAINS OF ORGANISMS ISOLATED FROM A CASE OF PNEUMOCOCCAL PNEUMONIA*

Date	Organism	Inhibiting concentration of aureomycin	
		Complete μg., ml.	Partial μg./ml.
June 27	<i>S. mitis</i>	6.3	3.1-1.6
27	<i>S. aureus</i>	3.1	1.6-0.8
27	<i>S. pyogenes</i>	0.8	0.4-0.1
27	<i>D. pneumoniae</i> , type 44	3.1	1.6-0.8
27	<i>D. pneumoniae</i> , type 44†	3.1	1.6
28	<i>S. mitis</i>	3.1	1.6
28	<i>S. aureus</i>	3.1	1.6-0.8
28	<i>D. pneumoniae</i> , type 44	3.1	1.6-0.4
28	<i>N. catarrhalis</i>	3.1	1.6-0.4
29	<i>S. mitis</i>	12.5	6.3-3.1
30	<i>S. mitis</i>	12.5	6.3-3.1

*Patient J. T., aureomycin, 0.5 gm. every 6 hours started on June 27 after sputum and blood were obtained for culture.

†Organism from blood culture, all others from sputum.

bites and ticks removed from her body 7 to 10 days before the onset of headache, severe stiff neck, and a rash. The spinal fluid was entirely negative. Symptomatic improvement was rapid and the patient was afebrile within 18 hours of the first dose. An initial blood culture was positive for group I meningococcus. Subsequent blood cultures were all negative.

Typhoid Fever, SALMONELLA Infections, and Colon Bacillemia. Of the 5 cases of typhoid fever, 4 were in children under 5 and one was in a man 60 years old. Treatment in the children was begun between the 5th and 10th day of illness. They all had positive blood cultures before aureomycin was started, but those taken after 1 or 2 days of therapy were all sterile. In the older patient, treatment was begun 4 days after a relapse with severe symptoms. Bacteremia in this patient persisted during 6 days of aureomycin therapy. The cultures of urine and stools in all of the 5 patients were negative for *S. typhosa* after the second day of treatment. Isolation of *Salmonella* from stools obtained during aureomycin treatment, however, was difficult, because the cultures were frequently overgrown with *Proteus*. Symptomatic improvement occurred soon after treatment was started in only 1 of these 5 patients and was more gradual in the others. Evaluation of the effect of the drug in these patients with typhoid fever was very difficult.

The illness in each of the 3 patients with the other *Salmonella* infections was very severe. Clinical and bacteriological improvement occurred in only one of them in relation to the onset of treatment. One of the 2 patients with *S. newport* infection and the patient with *S. suispestifer* bacteremia died, and the *Salmonella* persisted in these 2 patients throughout 3 and 11 days, respectively, of treatment with aureomycin, 0.5 gm. every

6 hours by mouth and 20 mg. intramuscularly twice a day. Autopsy revealed no focus in the former, and a large vertebral and psoas abscess which yielded *S. suispestifer* was found in the latter.

A patient with a persistent colon bacillus bacteremia and evidence of infection and possibly of malignancy in the abdomen showed temporary symptomatic improvement during treatment with aureomycin by mouth. The bacteremia recurred, however, and the patient died after an intestinal hemorrhage. There was no autopsy.

The relevant bacteriological findings in these 9 patients are summarized in TABLE 5.

TABLE 5

BACTERIOLOGICAL FINDINGS IN CASES OF *Salmonella* INFECTIONS AND IN COLON BACILLEMIA TREATED WITH AUREOMYCIN

Pa- tient	Organism	Sensi- tivity μg./ml.	Positive cultures in relation to aureomycin therapy				Clinical result
			Source	Before	During	After	
R.H.	<i>S. typhosa</i>	12.5	Blood	+	0	0	Good
			Stool	+	0	0	
J.H.	<i>S. typhosa</i>	12.5	Blood	+	—	—	Doubtful
			Stool	+	+ 1 day	0	
L.O.	<i>S. typhosa</i>	12.5	Blood	+	+ 1 day	—	Good ?
			Urine	0	0	—	
			Stool	0	0	—	
M.O.	<i>S. typhosa</i>	25	Blood	+	0	—	No effect?
			Urine	0	0	0	
			Stool	0	0	—	
A.L.	<i>S. typhosa</i>	6.3	Blood	+	+ 6 days	—	No effect?
			Urine	0	0	0	
			Stool	+	0	0	
M.E.	<i>S. typhosa</i> (chronic carrier)	12.5	Stool	+	+ 8th day	+	No effect
P.E.	<i>S. newport</i>	25	Duodenum	—	+ 30th day	—	
			Blood	0	0	0	Good
			Urine	0	0	—	
			Stool	+	0	0	
C.M.	<i>S. newport</i>	6.3	Blood	0	—	+ A [*]	Died
			Urine	+	—	+ A	
			Stool	+	+	+ A	
A.A.	<i>S. cholerae suis</i>	6.3	Blood	+	+	+ A ^{**}	Died
			Urine	0	—	—	
F.S.	<i>E. coli</i>	12.5	Blood	+	+	—	Died

* A = Autopsy culture.

** Same organism from vertebral and psoas abscess and from liver at autopsy.

Urinary Tract Infections. The cases of urinary tract infections that were treated with aureomycin were all severe. Seven of the 16 included in this study received a second course of a larger amount of aureomycin by mouth after an interval of about 2 weeks. All but one of these 16 patients were known to have urinary tract infections longer than one month, and in some instances, the infection had persisted for many years. In almost every instance there was a prolonged history of complicating factors such as urinary retention, frequent instrumentation, prostatectomy, urolithiasis, and chronic or recurrent acute cystitis and pyelonephritis. One

or more chemotherapeutic or antibiotic agents had previously been used in most of these patients without producing prolonged remissions.

Only one of the patients had acute pyelonephritis and had received no previous treatment. It is obvious, therefore, that the severest and the least favorable types of cases were chosen for an evaluation of the efficacy of aureomycin.

The predominant clinical symptoms in these cases were dysuria, poor sphincter control, nocturia, frequency and cloudy urine. The cultures obtained before aureomycin treatment usually revealed pure cultures of a single strain. In most patients who received a second course of aureomycin, organisms other than those initially present had appeared either during or after the first course.

Aureomycin was given, usually for 7 days, in doses of 0.5 gm. morning and evening. Those who were retreated received 0.5 gm. 4 times a day. The effect of still larger doses is now being studied. Only 4 of these patients were hospitalized during treatment.

In general, pyuria diminished markedly during therapy and disappeared entirely in about half of the cases while treatment was still being given. Symptomatic relief occurred at the same time in most instances; the burning on urination, nocturia and frequency all diminished. The urines became sterile during treatment in most of the patients. In some of the patients *Proteus* and *Pseudomonas*, and in 2 instances *E. coli*, either first appeared or persisted during aureomycin treatment. Infection recurred within a few days in most instances, with either the same or with different types of organisms. The most persistent organisms were the *Proteus* and, indeed, these seemed almost to flourish, but *Pseudomonas* also could not be eradicated.

Good results of a permanent nature were, therefore, infrequent in these cases of chronic urinary tract infections. Temporary relief, either slight or moderate, resulted in two-thirds of the cases.

A summary of the bacteriological findings, including the aureomycin sensitivity of the various strains isolated, is given in TABLE 6. It is seen

TABLE 6

ORGANISMS ISOLATED FROM CASES OF URINARY TRACT INFECTIONS TREATED WITH AUREOMYCIN

Organism	No. of pts.†	Sensi- tivity µg./ml	Number of cases		
			Before Rx	During Rx	After Rx
<i>Streptococcus mitis</i>	1	6.3	1	0	0
<i>Klebsiella pneumoniae</i>	1	50	1	1	
<i>Aerobacter aerogenes</i>	9 ¹	6.3-100	7	4	3*
<i>Escherichia coli</i>	9 ²	6.3-100	11	6	3
<i>Pseudomonas aeruginosa</i>	4 ¹	200-250	3	3	4
<i>Proteus vulgaris</i>	7 ⁴	125-250	4	8	9

* Two of these were the only strains of this organism resistant to 50 or 100 µg./ml.

† Superscripts indicate number of cases included in which the same organism was isolated in relation to 2 separate courses of therapy.

that the more resistant organisms, particularly *Proteus* and *Pseudomonas*, were the ones which persisted throughout treatment or appeared after treatment with aureomycin.

Lymphogranuloma Venereum. One patient with bilateral inguinal adenopathy of two weeks' duration and a positive Frei test was given a total of 40 gm. of aureomycin by mouth over a period of 20 days. The inguinal nodes on the right had enlarged progressively and became increasingly tender during the previous 2 weeks, and 2 days before treatment began to discharge foul pus which was bacteriologically sterile. During the aureomycin therapy, drainage and tenderness on that side diminished markedly within 4 days, and subsided almost completely after 2 weeks. The sinus tract persisted and looked clean and the swelling regressed considerably. The nodes on the left increased in size and in tenderness during the treatment; they became fluctuant by the 11th day, and were aspirated on that day and again on the 18th day. There was progressive diminution in the swelling and tenderness, but no sinus tract developed. After aureomycin was discontinued, sulfadiazine was given for 5 days without evidence of any further improvement.

Measles. A 4-year-old child was given 1 gm. of aureomycin during the third day of fever and first day of rash. Fever persisted at 104° to 105° and the drug was discontinued because of looseness of the bowels and prostration. The course in this child was identical to that of his brother, who was not treated.

Non-specific Urethritis. Two patients were referred by Dr. Howard M. Trafton, because of persistent urethritis in spite of repeated therapy with penicillin, sulfonamides, and streptomycin. Cultures were negative for *Gonococcus* and in one of the patients a pleuropneumonia-like organism was isolated. In the latter patient, a total of 9 gm. of aureomycin was given by mouth over a period of 6 days. The patient was symptomatically improved within 3 days and the discharge stopped and only occasional cells could be seen in the prostatic smears. Cultures were negative after treatment. The other patient was given 11 grams in 5 days. He became asymptomatic quite rapidly. The discharge began to clear before the treatment was ended and was completely gone within 11 days.

Toxicity. Significant toxic effects were notably absent. Doses of 250 mg. or less, twice a day, were used in adults at first, but these have been increased progressively until recently an initial dose of 4 gm. with daily doses of 2 gm. have been used in many cases. The treatment has been maintained in most patients, except those with gonorrhea, for 1 to 2 weeks on this dosage and as long as one month on a dose of one gram a day. Minor symptoms were more frequent with the higher doses.

The commonest complaint during oral therapy with aureomycin was looseness of the bowels with frequent, bulky, and soft stools. True diarrhea was uncommon. Normal bowel action again occurred when the drug

was discontinued or when the dose was temporarily decreased. Nausea and, occasionally, vomiting occurred after one or more doses in a few patients but could not always be related to the treatment. In some of the patients with cystitis, a disagreeable sensation of as something "drawing" or "squirming" in the pelvis was noted. This may be related to the high acidity of the urine during ingestion of large doses of aureomycin. Other symptoms were not noted.

The patients treated in the hospital were followed carefully with frequent blood counts and in some of them liver function tests were done. Anemia ascribable to the drug did not occur. No depression of the cells in the granulocytic series was observed. There was no evidence of renal irritation as manifested by rising non-protein nitrogen levels in the blood, albuminuria or appearance of cellular elements in the urine. Jaundice did not develop in any of the patients, nor was there any evidence of liver impairment developing after treatment was started. No rashes or fever ascribable to the drug were observed. In a 3-year-old patient with typhoid fever, fever recurred and increased after the temperature had almost reached normal. The patient remained otherwise well and was, therefore, suspected of having drug fever. A single 200 mg. dose given after the child again was afebrile for a few days produced no fever or symptoms, so that the relation of the drug to the persistent fever could not be evaluated.

Intramuscular doses were used in only 3 patients. In the first of these patients, the injection of 20 mg. caused considerable local pain, but in the others, who were severely ill and were given similar amounts repeatedly for several days, there was no apparent discomfort. Sterile saline was used as a diluent in these latter cases.

Bacterial Resistance. Studies of the sensitivity of the organisms which were obtained in the same patients before, during, and after treatment with aureomycin yielded no evidence of the development of bacterial resistance to aureomycin *in vivo*.

Conclusions

Aureomycin by mouth in the doses used exerted a definite beneficial effect in the coccal infections. The results in the cases of gonorrhea, however, were inferior to those obtained with adequate amounts of penicillin. In the urinary tract infections, when the bacteriological findings and the type of cases chosen for treatment are taken into consideration, the results were favorable and may be comparable with, or possibly superior to those resulting from the use of streptomycin in similar cases. Infections with *Proteus vulgaris* and *Pseudomonas aeruginosa*, however, were not benefited by aureomycin. The results in the cases of typhoid fever and of severe *Salmonella* infections were equivocal. The only evidence of toxicity was the occurrence of copious, loose bowel movements when large doses were given by mouth. Resistant organisms did not develop during treatment with this antibiotic.

References

1. PAINE, T. F., H. S. COLLINS, & M. FINLAND. 1948. Laboratory studies with aureomycin. *Ann. N. Y. Acad. Sci.* **51**(2): 228-230.
2. COLLINS, H. S., E. B. WELLS, T. F. PAINE & M. FINLAND. Urinary excretion of aureomycin. Preliminary report. *Proc. Soc. Exp. Biol. & Med.* (In press.)

STUDIES ON SERUM CONCENTRATIONS IN HUMANS AND PRELIMINARY OBSERVATIONS ON THE TREATMENT OF HUMAN INFECTIONS WITH AUREOMYCIN*

By HARRY F. DOWLING, MARK H. LEPPER, LEWIS K. SWEET,
AND ROBERT L. BRICKHOUSE

*George Washington University Medical Division and Infectious Disease Service
Gallinger Municipal Hospital, and Departments of Medicine and Pediatrics,
George Washington University, Washington, D. C.*

AUREOMYCIN,* a new antibiotic, which has been found to be effective *in vitro* and in animals against a variety of microorganisms, has been employed by us in human infections. In the present paper, we are reporting the results of preliminary studies, which include: (1) The determination of the concentration of the antibiotic in the blood after oral and intramuscular doses, (2) observation of patients for toxic reactions after single and multiple doses, and (3) clinical tests in a few patients.

Determination of Blood Aureomycin Concentrations. Attempts were made to determine the amount of aureomycin present in the serum with the use of the tube dilution method, making readings after twenty-four hours as in the determination of penicillin and streptomycin. It soon became apparent that the inhibiting effect of aureomycin upon bacteria was seldom present for as long as twenty-four hours. Incubation at 37° C. of 1 microgram per cc. or less of aureomycin, in tryptose phosphate broth, thioglycollate medium, or human serum, showed that only one-sixteenth to one-thirty-second of the original activity remained after eight hours. Neither the presence or absence of organisms, nor culturing under oil, made any difference in these results.

The effect of the antibiotic upon the growth of a strain of *Bacillus cereus*† is shown in FIGURE 1. This test was performed by placing in a series of tubes 3 cc. of a 1:10,000 dilution in tryptose phosphate broth of a 16-hour culture of *Bacillus cereus*. To one tube, 1 cc. of sterile broth was added and to other tubes, aureomycin in broth, to make a final dilution of 10, 2.5, 0.625, and 0.039 micrograms, respectively. The tubes were incubated at 37° C. and at 1, 2, 4, 6, 8, 12, and 24 hours after the start of incubation, 0.1 cc. of the mixture was withdrawn from each tube, diluted 1:1000 or 1:10,000, and plated after mixing with nutrient agar. The number of colonies on each plate were counted after twenty-four hours' incubation.

* Supplied by the Lederle Laboratories Division, American Cyanamid Company.

† This strain was known as *Bacillus* No. 5 and was supplied by Mr. A. C. Dornbush.

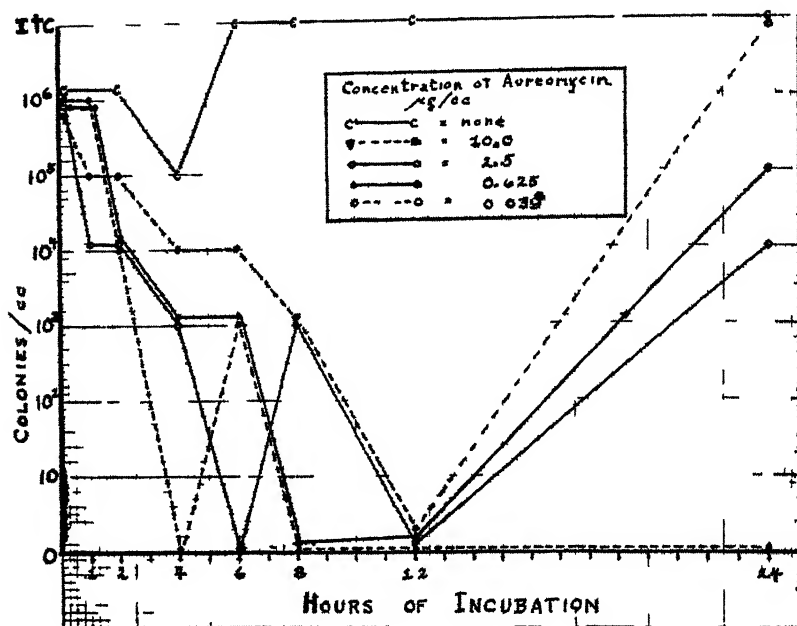


FIGURE 1 The effect of varying concentrations of aureomycin upon the growth of strain B. H. 1704

FIGURE 1 shows the result of a typical experiment. Plates taken from the control tube (without aureomycin) contained colonies too numerous to count (ITC) from the sixth hour on. The mixture containing 10 micrograms per cc of aureomycin inhibited growth of all organisms from the eighth hour on while those with smaller amounts of aureomycin showed inhibition of growth up to the twelfth hour and growth again by the twenty-four hour. In repeated experiments, the period of maximum depression of the bacterial count was found to be between the fourth and the twelfth hours.

From our data, therefore, it would appear that when a method of assay is used in which readings are made at twenty-four hours, only those concentrations are detected which are high enough initially to have complete bactericidal action before the aureomycin is inactivated or which are high enough initially to retain some bacteriostatic action for twenty-four hours.

Since aureomycin is merely bacteriostatic in low concentrations, in determining serum concentrations in patients, we followed the method of Dornbush,¹ namely, incubation for four hours in a water bath. The only modification of his method was the use of tryptose phosphate broth.

TABLES 1, 2, and 3 show the serum concentrations of aureomycin after the administration of single doses intramuscularly or by mouth. After

TABLE 1

NUMBER OF SPECIMENS AT EACH BLOOD CONCENTRATION AFTER
ADMINISTRATION OF 100 MG I.V.

Concentration	Hours						
	1	3	6	8	12	18	24
2.0		1					
1.0	2			1			
0.5		2					
0.25	3	1	1				
0.125					2		
0.06		1	1	1	3	1	1
0					1		5
Total	5	5	5	2	6	1	6

the intramuscular injection of 100 mg. to adults (TABLE 1) the peak concentration of the antibiotic was reached at about the third hour. Detectable concentrations were present in the blood serum at the twelfth hour in nearly all patients and in the case of one patient as long as twenty-four hours after injection.

When 700 mg. were given by mouth (TABLE 2) the peak concentration in the serum was reached around the sixth hour and the antibiotic was

TABLE 2

NUMBER OF SPECIMENS AT EACH BLOOD CONCENTRATION AFTER
ADMINISTRATION OF 700 MG. P.O.

Concentration	Hours						
	1	3	6	8	12	18	24
1.0			1				
2.0		2		2			
1.0	2	2	3	1	1		
0.5	3		2				
0.25		3		2			
0.125	3	1	1		2		
0.06					2	1	
0						2	3
Total	8	8	9	5	5	3	3

TABLE 3

NUMBER OF SPECIMENS AT EACH BLOOD CONCENTRATION AFTER
ADMINISTRATION OF 300 MG. P.O.

Concentration	Hours of administration					
	1	3	6	12	18	24
1.0		2	2			
0.5	1	2	2			
0.25	2	1	1	4		
0.125	1					
0.06				1		3
0						2
Total	4	5	5	5		5

detectable in the sera of all patients tested at the twelfth hour. After the oral administration of 300 mg., the peak levels were not as high as with 700 mg., but aureomycin was still detectable in the serum of all the patients at twelve hours.

Five patients who received 10 mg. per kilogram of body weight at six-hour intervals showed no appreciable difference in serum concentrations on the third to the seventh day of treatment compared with the first day of treatment. Although the optimal dosage schedule still needs to be worked out, it appears that oral doses of 5 to 10 mg. per kilogram at six-hour intervals are sufficient to assure that measurable amounts of aureomycin will be consistently present in the blood.

Toxic Reactions. We have observed no untoward reactions to aureomycin, except pain at the site of intramuscular injection. This pain was present in all of our patients who were given intramuscular injections. The dose varied from 100 mg. to 1400 mg. The pain was an immediate stinging or aching sensation at the site of injection which lasted from 15 minutes to 2 hours. It was more intense and lasted longer when the larger doses were given. In no instance was there any delayed reaction or any generalized effect from the injections. We have given aureomycin intramuscularly to patients in doses of 5 to 1400 mg. and orally in doses of 100 to 700 mg. as shown in TABLES 4 and 5

TABLE 4
NUMBER OF PATIENTS TREATED INTRAMUSCULARLY

<i>Dose (mg.)</i>	<i>Single doses</i>	<i>Pt.</i>	<i>Days</i>
100		7	
300		2	
350		3	
700		2	
1400		1	
	<i>Multiple (daily) doses</i>		
5, 10 & 20		1	3
40, 80, 160, 320		1	4
700, 300, 350		1	3

TABLE 5
NUMBER OF PATIENTS TREATED ORALLY

<i>Dose (mg.)</i>	<i>Single doses</i>	<i>Pt.</i>	<i>Days</i>
700		11	
350		1	
300		4	
	<i>Multiple doses</i>		
700 q6h		2	7
700 q6h		1	3
300 q6h		1	7
250 q6h		1	6

Observations on Patients Treated. Four patients with typhoid fever were treated. In two, the treatment was started during the second week and in two, during the third week of the disease. The patients received 10 mg. per kilogram of body weight orally every four hours. No effect upon the clinical course was observed. Blood cultures remained positive during treatment and, in one patient, the sensitivity of the organism changed from 0.6 micrograms per cc. to 20 micrograms per cc. during six days of therapy. *E. typhi* could not be cultured from the stools of any of the patients however, after therapy was started.

Two patients with Rocky Mountain spotted fever were treated. The first, an 11-year-old boy, was given 300 mg. every six hours. The temperature began to subside within twenty-four hours and was within normal limits in sixty hours. The rash began to disappear with the drop in temperature and was entirely gone in twenty-four hours. Aureomycin therapy was continued for five days and recovery was uneventful.

The second patient, a 43-year-old man, had been under treatment with sodium para-amino-benzoate for 10 days without effect. He was disoriented and had continuous fever of 104° to 105° F. Other medication was discontinued and aureomycin treatment started in doses of 700 mg. every six hours. Fever and toxicity began to subside within twelve hours. The temperature reached normal, forty-eight hours after the start of therapy, and remained within normal limits thereafter. Aureomycin treatment was given for five days altogether and recovery was uneventful.

Summary

1. Assayable serum concentrations of Aureomycin may be obtained when the antibiotic is given by either the intramuscular or the oral route. The former is somewhat painful so that the oral route is recommended as a routine. Doses at six-hour intervals have been found to be satisfactory, although eight- or twelve-hour intervals may be sufficient.

2. Four patients with typhoid fever showed no clinical or bacteriological response when treated with aureomycin, while two patients with Rocky Mountain spotted fever appeared to recover dramatically after the institution of aureomycin therapy.

Reference

1. DORNBUSH, A. C., & E. J. PELCAK. 1948. The determination of aureomycin in serum and other body fluids. *Ann. N. Y. Acad. Sci.* 51(2): 218-220.

USE OF AUREOMYCIN ON SOME EXPERIMENTAL INFECTIONS IN ANIMALS

By PAUL A. LITTLE

Lederle Laboratories Division, American Cyanamid Company Pearl River, N. Y.

AUREOMYCIN has been tested in mice against experimental infections due to *Eberthella typhosa* and to *Erysipelothrix rhusiopathiae*. Baby chicks have been used to test Aureomycin against experimental infections due to organisms which are responsible for highly fatal, naturally occurring diseases of chicks. These organisms are *Pasteurella multocida* and *Shigella gallinarum*. We have found Aureomycin effective against the infections produced when each of these organisms was inoculated into the experimental animals.

TABLE 1 summarizes the methods by which the different infecting agents were prepared for animal inoculation.

TABLE 1
METHODS OF PREPARING INFECTING AGENTS
FOR ANIMAL INOCULATION

ORGANISM	TIME OF INCUBATION OF CULTURE (37° C)	SIZE OF INOCULUM (0.5 ML)
E. TYPHOSA	24 HRS	SUSPENSION*
E. RHUSIOPATHIAE	16 HRS	10 ⁻⁴ DILUTION
P. MULTOCIDA	6 HRS	10 ⁻² DILUTION
S. GALLINARUM	6 HRS	10 ⁻⁶ DILUTION
*SUSPENSION FROM BLOOD AGAR PLATE WAS STANDARDIZED TO CONTAIN ABOUT 1 BILLION ORGANISMS PER DOSE. ALL DILUTIONS WERE MADE IN 2 PER CENT PEPTONE SOLUTION.		

Eberthella typhosa is of interest as a representative of gram-negative organisms of the typhoid-paratyphoid group. The fact that the organism is not naturally virulent for mice necessitated inoculating suspensions standardized to contain about 1 billion organisms per 0.5 milliliter dose. This amount of culture was obtained by suspending the growth from a blood agar plate in 2 per cent peptone solution. Given intra-abdominally, the dose killed mice in 24 hours.

Erysipelothrix rhusiopathiae is the cause of the economically important disease, swine erysipelas, and is naturally infectious to many animals including man. Our demonstration of the ability of Aureomycin to pro-

fect mice against *E. rhusiopathiae* serves to differentiate the mode of action of this antibiotic from that of Polymyxin which is effective against *E. typhosa* but not against *E. rhusiopathiae*. Since the organism is highly virulent for mice, we used 0.5 milliliter of a 10^{-4} dilution of culture grown for 16 hours in bile-enriched yeast-extract broth. The organism has a peculiar nutritional requirement for fatty acids furnished by bile. The inoculum used contained at least 1000 minimum lethal doses and killed mice in 48 to 72 hours.

Pasteurella multocida is of interest as a representative of the group of gram-negative organisms which includes *Pasteurella pestis*. Besides causing fowl cholera, this organism produces rapid and highly fatal infections in all species of domestic and wild animals, but is innocuous to man. We inoculated one-day-old chicks intra-abdominally with 0.5 milliliter of a 10^{-2} dilution of culture grown for 6 hours in yeast-extract broth. This amount of inoculum contained at least 100,000 minimum lethal doses and killed chicks in 24 to 48 hours.

Since it is unusual to employ chicks as laboratory animals for bacteriological studies, it may be of interest that they weigh about 40 grams when one day old and may be injected intra-abdominally as many as six times a day without fatality. We used a pure strain of New Hampshire Red chicks furnished by a local hatchery.

Shigella gallinarum, the cause of fowl typhoid, is of interest because it produces typical symptoms of typhoid in the chick. We used 0.5 milliliter of a 10^{-6} dilution of culture grown for 6 hours in yeast-extract broth. This amount of inoculum, injected intra-abdominally, contained at least 100 minimum lethal doses and killed chicks in 72 to 120 hours. The culture was also virulent when given by mouth.

Experimental Results

In testing Aureomycin against *E. typhosa* in mice, we have studied the effect of starting treatment $\frac{1}{2}$ hour before inoculating the culture, and the effect of delaying treatment until $\frac{1}{2}$ hour after culture. In both experiments the doses of Aureomycin were divided and the half dose was given intra-abdominally at 9 A.M. and 5 P.M. for three days.

FIGURE 1 shows the endpoint of activity of Aureomycin when treatment was started $\frac{1}{2}$ hour before culture. One milligram per kilogram per day did not protect; two milligrams gave 50 per cent; four milligrams gave 80 per cent; eight milligrams gave 100 per cent protection. The actual amounts of Aureomycin administered may be of interest. To obtain 50 per cent protection, doses of 0.02 milligram were given at 9 A.M. and 5 P.M. for three days. To obtain 100 per cent protection, doses of 0.08 milligram were given.

FIGURE 2 shows the effect of delaying treatment until $\frac{1}{2}$ hour after inoculating the culture. The range of from 5 to 40 milligrams per kilogram per day includes the smallest amount of Aureomycin expected to give complete protection on the basis of the results of the previous ex-

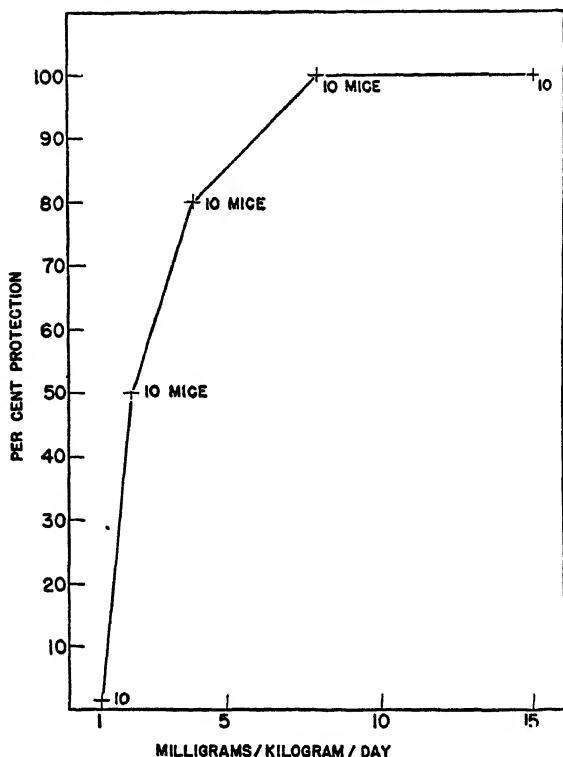


FIGURE 1 Action of Aureomycin against *E. typhosa*. Treatment started $\frac{1}{2}$ hour before culture.

periment. This amount of Aureomycin gave 90 per cent protection under the conditions of the test. To obtain this result doses of 0.05 milligram were given at 9 A.M. and 5 P.M. for three days. Only 70 and 80 per cent protection were obtained with 10 and 20 milligrams per kilogram. Complete protection was obtained with 40 milligrams. The curve does not show a different endpoint of activity from that demonstrated in our first experiment. Since 100 per cent of mice receiving 40 milligrams per kilogram lived, the curve does not show a loss of mice which might be attributed to toxicity of the antibiotic. The losses of mice incurred in this experiment appear to be due solely to the severity of the test conditions. In other words, the results illustrate the well-known importance of early treatment.

FIGURE 3 shows the endpoint of activity of Aureomycin when used to treat mice infected with *E. rhusiopathiae*. The doses were divided and the half dose given at 9 A.M. and 5 P.M. for three days, starting $\frac{1}{2}$ hour before culture. Four milligrams per kilogram per day did not protect; eight milligrams gave 20 per cent; fifteen milligrams gave 80 per cent; thirty milligrams gave complete protection. To obtain 50 per cent protec-

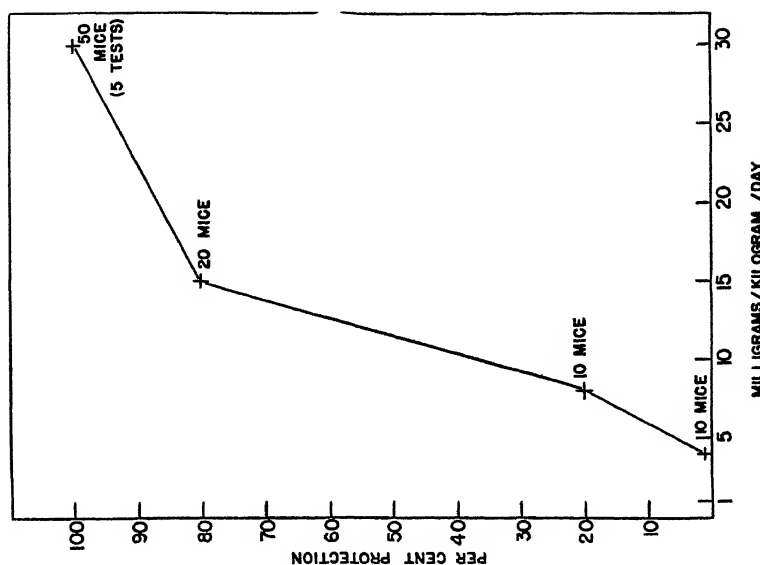


FIGURE 2. Action of Aureomycin against *E. typhosa*. Treatment started 1/2 hour before culture.

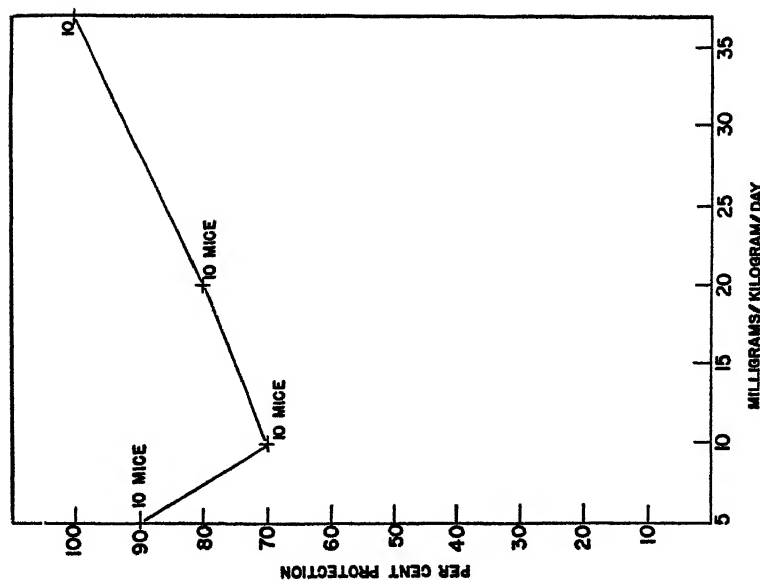


FIGURE 3. Action of Aureomycin against *E. rhusiopathiae*. Treatment started 1/2 hour after culture.

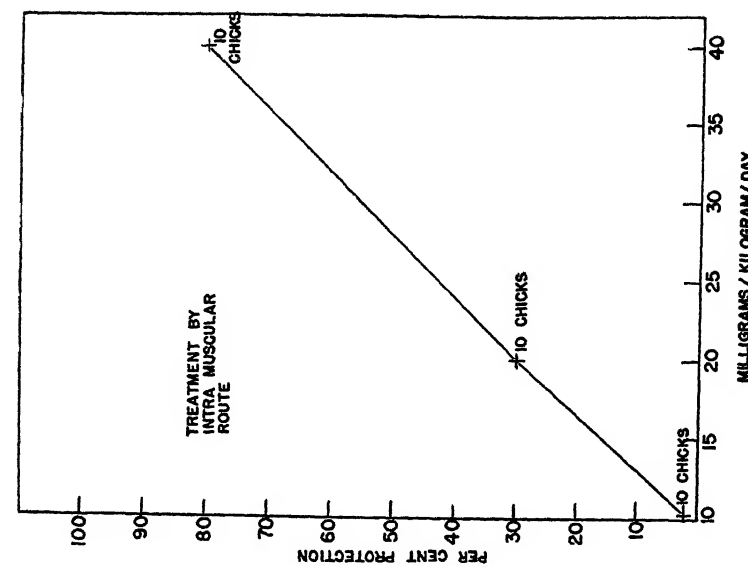


FIGURE 3. Action of Aureomycin against *P. multocida*. Treatment started 1 hour after culture.

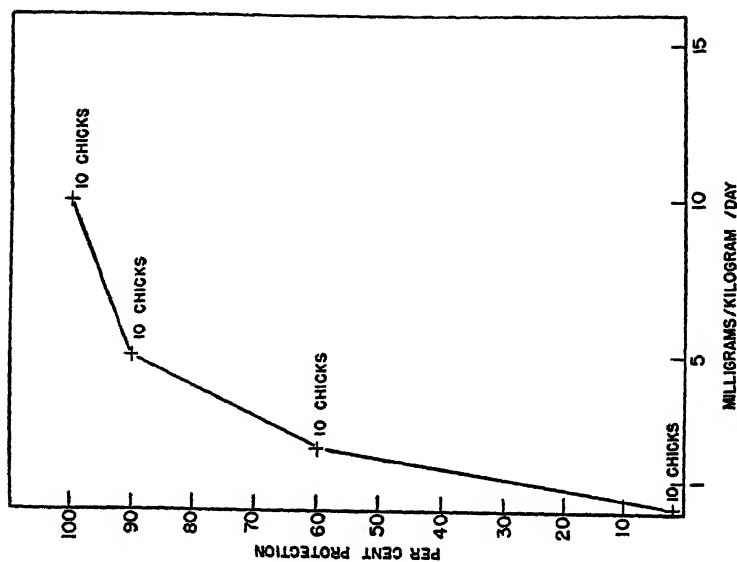


FIGURE 4. Action of Aureomycin against *P. multocida*. Treatment started 1 hour before culture.

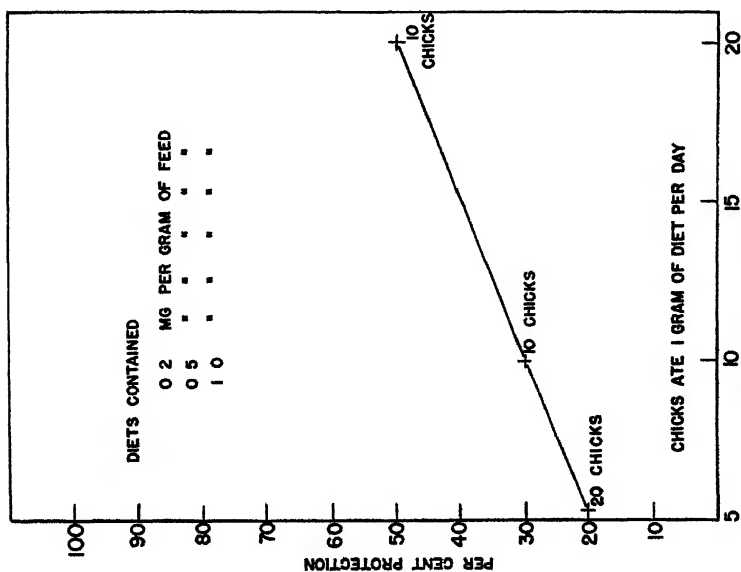


FIGURE 6 Action of Aureomycin against *P. multocida* Effect of orally administered Aureomycin

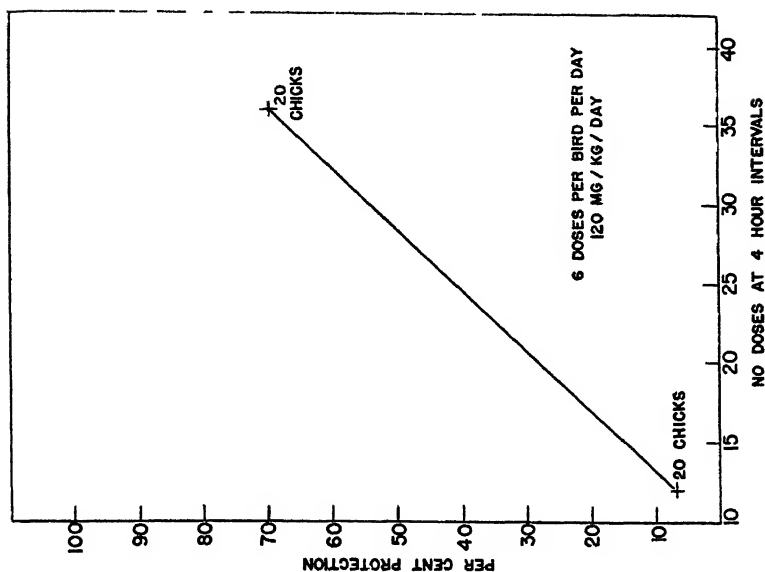


FIGURE 7 Action of Aureomycin against *P. multocida* Effect of treating at 4 hour intervals using doses of 1 mg

tion would require 11.5 milligrams as compared to 2 milligrams required in our tests with *E. typhosa*. Apparently 5 to 6 times as much Aureomycin is required in treating infections due to *E. rhusiopathiae* as in treating infections due to gram-negative bacteria. It is of interest that penicillin is also effective against *E. rhusiopathiae*.

FIGURE 4 shows the endpoint of activity of Aureomycin when used to treat one-day old chicks infected with *P. multocida*. In this experiment, the chicks received intra-abdominal injections of Aureomycin once daily for four days, starting $\frac{1}{2}$ hour before culture. Two milligrams per kilogram gave 60 per cent protection; five milligrams gave 90 per cent; and ten milligrams gave complete protection. Even though the chicks received only one treatment per day, the amount of Aureomycin required to give 50 per cent protection corresponds closely with the result of our tests with *E. typhosa*. In these experiments, two milligrams per kilogram per day appears to be the smallest effect dosage of Aureomycin.

FIGURE 5 shows the results of an experiment in which Aureomycin was administered by the intramuscular route and treatment was delayed until $\frac{1}{2}$ hour after culture. The doses were divided and the half dose given in the right leg at 9 A.M. and 5 P.M. for four days. Ten milligrams per kilogram failed to protect; twenty milligrams gave 30 per cent protection; and forty milligrams gave 80 per cent protection against *P. multocida*. This experiment answers a question which might be raised with respect to the previous experiments as to whether or not the antibiotic is active when given by a different route from that used in inoculating the culture.

FIGURE 6 shows the effect of orally administered Aureomycin in treating chicks infected with *P. multocida*. The antibiotic was mixed with a commercial chick starter and fed *ad libitum* for four days, starting 24 hours before culture. The chicks were one day old when introduced to the diet and two days old when inoculated with culture. At this age, chicks eat about 1 gram of feed per day. This approximation was used in determining the position of the curve with respect to milligrams of antibiotic per kilogram body weight per day. Three preparations of feed were used in this experiment. One contained 0.2 milligram of Aureomycin per gram; one contained 0.5 milligram; and one contained 1 milligram. In preparing 1 milligram of Aureomycin per gram of feed, we dissolved 500 milligrams of antibiotic in 5 milliliters of distilled water and added this to 500 grams of commercial chick starter. The container was cleaned by the absorbing action of the dry feed.

The diet containing the smallest amount of antibiotic had some protective effect since the culture regularly killed 100 per cent of untreated controls. The diet containing 1 milligram of antibiotic per gram protected 50 per cent of chicks. These chicks appear to have consumed 20 milligrams of antibiotic per kilogram body weight per day.

FIGURE 7 shows the action of Aureomycin against *S. gallinarum* in chicks. In preliminary experiments we had found that this organism was

resistant to twice daily doses of the antibiotic. In FIGURE 7 we have compared the effect of 12 doses of 1 milligram, given at 4-hour intervals, and the effect of 36 doses. The more intensive treatment protected 70 per cent of chicks

TREATMENT OF EXPERIMENTAL INFECTIONS WITH AUREOMYCIN*†

By MORTON S. BRYER, EMANUEL B. SCHOENBACH,
ELEANOR A. BLISS, AND CAROLINE A. CHANDLER

*Department of Preventive Medicine,
The Johns Hopkins University School of Medicine,
Baltimore, Maryland*

AUREOMYCIN is an antibiotic derived from a strain of *Streptomyces*. It was supplied as a yellow crystalline hydrochloride salt which is soluble in distilled water but somewhat less soluble in normal saline solution. These solutions are acid (pH 4.5) and their activity deteriorates rapidly with increasing pH at room temperature.

Much of the *in vitro* work, including the methods of biological assay utilized in securing body fluid levels, has been described by Dr. Chandler and Dr. Bliss.‡ It is the purpose of this presentation to describe some of the pharmacological studies and experimental infections which have been performed.

Toxicity. The acute toxicity of aureomycin was determined in mice and dogs. All mice survived a single intravenous injection of 50 mg./kg. Six out of 7 mice died after receiving 100 mg. kg. Dogs developed transient hyperpnea, generalized weakness, and anorexia following rapid intravenous injection of 50 and 100 mg. kg. Injection of 150 mg./kg. was followed by grunting respirations, tremors, generalized paresis, and somnolence. Death occurred 6 hours after administration of this dose. At autopsy, hemoglobinuria was noted in the bladder urine. Since all intravenous injections were with 10 per cent aureomycin in distilled water, the acidity of such a large dose may well have caused hemolysis.

The majority of mice survived, for a period of 7 days, single subcutaneous injections up to 3000 mg. kg. Only occasional deaths occurred in animals receiving 1000 to 3000 mg. kg. All mice died when 4000 mg. kg. was administered subcutaneously. Death was preceded by hyperpnea, anorexia, tremors, ataxia, and paresis.

Rats tolerated subcutaneous injections of 50 mg. kg. once daily for 8 days with only slight weight loss and mild local inflammatory reaction at the injection site. A dog which received 20 mg. kg. intramuscularly twice a day for 9 days had marked persistent anorexia and progressive weight loss. Induration, necrosis, and fluctuation of the injected tissues was noted. The rats and dogs were injected with aureomycin at a concentra-

* Aureomycin used in these experiments was supplied by the Lederle Laboratories Division, American Cyanamid Company.

† These investigations were supported by grants received from Abbott Laboratories, Eli Lilly and Company; Lederle Laboratories Division, American Cyanamid Company, Parke, Davis and Company; and the Upjohn Company.

‡ See their paper in this monograph.

TABLE 1
ACUTE TOXICITY OF AUREOMYCIN (HCL)

Mode of administration	Dose mg./kg.	Animal		Death in 2 days	Survival 7 days
		Type	No		
Intravenous (rapid)	50	Mice	3	0	100%
	100	Mice	7	6	14%
	250	Mice	1	1	0
	500	Mice	1	1	0
	50	Dog	1	0	1
	100	Dog	1	0	1
	150	Dog	1	1	0
Subcutaneous	500	Mice	10	0	100%
	1000	Mice	14	1	93%
	1500	Mice	10	1	90%
	2000	Mice	17	2	88%
	3000	Mice	6	2	77%
	4000	Mice	5	5	0

Signs of intoxication—Rapid respirations, tremors, ataxia, and paresis

tion of 2.0 to 10.0 per cent in 1.0 per cent procaine hydrochloride in distilled water.

Three drops of 0.25 per cent and 1.0 per cent aureomycin borate solution were instilled locally into the conjunctival sac of a rabbit. No immediate or delayed local reaction was noted.

TABLE 2
AUREOMYCIN (HCL) CHRONIC TOXICITY

Dosage schedule					
Animal	Daily dose mg./k.	No. doses per day	No. days administered	Total dose mg.	Observed for 10 days Remarks*
Rat 1	50 S.C.	1	8	112	12 gm. wt. loss
2	50 S.C.	1	8	130	10 gm. wt. loss
3	Control	—	—	—	3 gm. wt. gain
4	Control	—	—	—	20 gm. wt. gain
Dog 1	40 I.M.	2	9	2,280	1200 gm. wt. loss. Marked reaction at site of injections
2	40 I.M.	1	1	160	Well tolerated
Rabbit 1	20 I.M.	1	1	70	Well tolerated
2	Locally in the eye (borate)	0.25%	1	0.5	No reaction
		1.0%	1	2.0	No reaction

S.C. = Subcutaneously.

I.M. = Intramuscularly.

Locally—3 drops of the borate was instilled locally in the eye.

* Animals were observed for changes in weight and hemoglobin. Urine was examined for proteinuria, volume, specific gravity, and microscopic elements.

TABLE 3

ASSAY OF BODY FLUIDS OF ANIMALS INJECTED INTRAMUSCULARLY WITH AUREOMYCIN

Animal	Dosage schedule	Day of adminis- tration	Interval post last injection	Levels of aureomycin gamma/cc.	
				Serum	Sp. fl. urine
Rabbit	20 mg. kg. single dose	1	0	0	
			30 Min.	1.25	
			1 Hour	1.25	
			2.5 Hrs.	0	
			3.5 Hours	0	
			6 Hours	0	
Dog 1	40 mg./kg. single dose	1	0	0	0
			15 Min.	1.2	0
			30 Min.	Trace	0
			1 Hour	0	0
			2 Hours	0	0
2	20 mg. kg. B.I.D. for 10 days	1	0	0	
			1 Hour	Trace	
		2	1 Hour	0	
		4	15 Hours	0	
			1 Hour	1.25	
		6	15 Hours	Trace	
			10 Hours	Trace	
		10	5 Min.	0.6	0
			30 Min.	0.6	0
			2.5 Hours	0.3	0

> 64

Autopsies were performed on mice, rats, and dogs which had received large (up to 3000 mg./kg.) and repeated doses (total up to 400 mg./kg. over a period of 9 days) of aureomycin. No gross or microscopic* abnormalities in the liver, kidney, heart, or bone marrow were noted in any of the animals. However, necrosis of tissue occurred at the local injection sites. Mice which had received aureomycin by intubation showed evidence of local erosions of the gastric mucosa. These were colored yellow by the drug.

Levels in Body Fluids. Blood serum levels were obtained following single intramuscular injections of 20 mg./kg. in rabbits and 40 mg./kg. in dogs. Levels of 1.25 micrograms/ml. were recorded in the serum 15 minutes to one hour following the injections. No significant serum concentrations could be measured when serum was obtained more than one hour after injection of the drug. A dog received 20 mg./kg. twice a day for 10 days. No significant level was determined in the serum obtained more than 2.5 hours after the last injection. Blood drawn at regular intervals one hour after injection during the period of administration of aureomycin contained 0.3 to 1.25 micrograms/ml. of serum. A random specimen of urine obtained on the 10th day contained 64 micrograms of aureomycin per ml. The antibiotic was not detected in the spinal fluid after parenteral administration.

* Microscopic sections were prepared and their interpretation corroborated by Dr. Tobias Weinberg, Pathologist, Sinai Hospital, Baltimore, Maryland.

Therapeutic Effectiveness: BLOOD CULTURES. Mice were infected intraperitoneally with 10,000 LD's of *Pneumococcus* type I (SVI) and *Streptococcus hemolyticus* beta (C203). Complete clearing of the blood stream infection occurred when a loopful of tail blood was cultured 18 hours after subcutaneous treatment with three injections of aureomycin or penicillin G 5 mg./kg. each. The control mice all died with positive blood cultures in this period. Penicillin G protected a higher proportion of the mice from death when observed over a 7 day period. Only partial clearing of the *K. pneumoniae* A from the bloodstream was obtained with a dosage schedule of 10 mg./kg. three times daily. Polymyxin gave complete clearing of the bacteremia, and streptomycin almost complete clearing. The dosage of polymyxin and streptomycin employed was only one-tenth that of aureomycin used above. None of the mice which had received aureomycin survived a 7-day period, while 40 per cent survived when treated with the smaller dosages of polymyxin or streptomycin.

TABLE 4
BLOOD CULTURES IN MICE* TREATED WITH AUREOMYCIN,
POLYMYXIN, OR STREPTOMYCIN

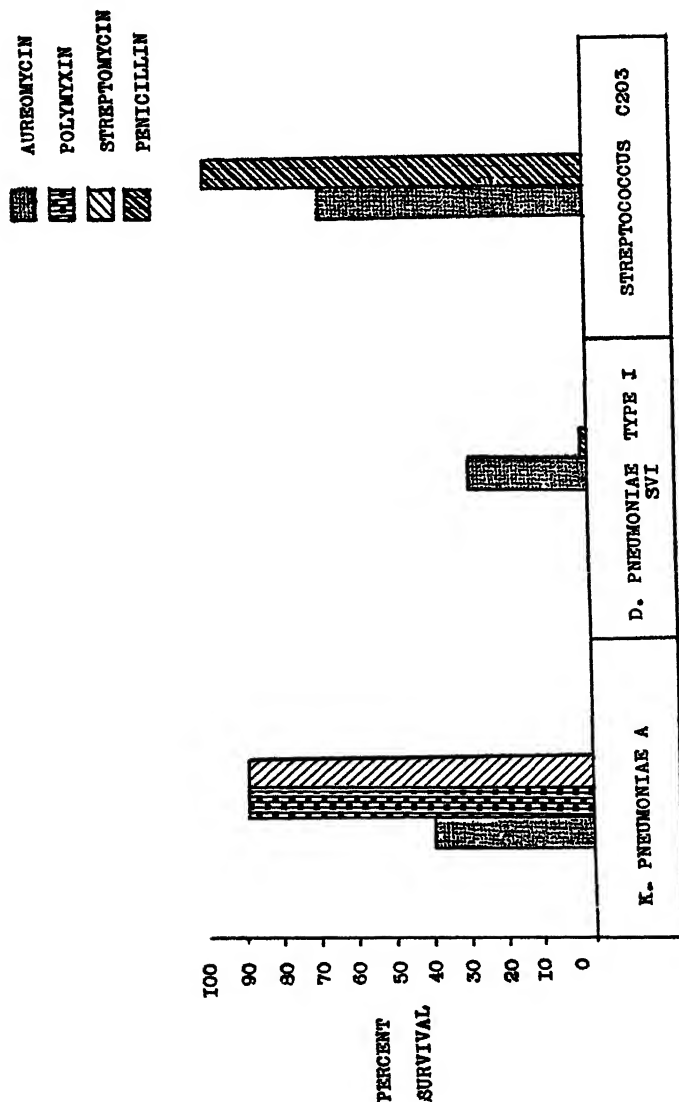
(Hemolytic streptococcus C203, *Pneumococcus* SVI, or *K. pneumoniae* infection)

No. mice	Organism	Drug	Dose mg./ kg.	Time after infection	Tail blood cultures Mouse					Per cent sur- vival 7 days
					1	2	3	4	5	
	<i>K. pneu- moniae</i> 10,000 LD'S	Untreated mice	0	2 Hrs.	+	+	+	+	+	
10		Aureomycin	30	20 Hrs.	0	+	0	+	+	0
10		Polymyxin	3		0	0	0	0	0	40
10		Streptomycin	3		0	0	+	0	0	40
16		Control	0		+	+	+	+	+	0
	Hemolytic streptococcus C203	Untreated mice	0	2 Hrs.	0	0	0	0	+	
10	10,000 LD'S	Aureomycin	15	20 Hrs.	0	0	0	0	0	50
10		Penicillin G	15		0	0	0	0	0	80
16		Control	0		+	+	+	+	+	0
	Type I <i>Pneu- mococcus</i> SVI	Untreated mice	0	2 Hrs.	+	+	+	+	+	
10	10,000 LD'S	Aureomycin	15	20 Hrs.	0	0	0	0	0	90
10		Penicillin G	15		0	0	0	0	0	100
16		Control	0		+	+	+	+	+	0

* Albino Swiss mice 18-22 gm. Initial dose of drug 2 hours post intraperitoneal infection and repeated T.I.D. for 3 days. Treatment subcutaneously.
† Dead at 20 hours post infection.

Subcutaneous Injection. A single subcutaneous injection of 50 mg./kg. of aureomycin administered immediately after infection, protected 40 per cent of mice infected intraperitoneally with 1000 LD's of *K. pneu-*

moniae A, 30 per cent of those infected with *Pneumococcus* type 1 (SVI), and 60 per cent of *Streptococcus hemolyticus* beta (C203) infected mice. Both polymyxin and streptomycin at the same dosage resulted in 90 per cent of mice surviving the *K. pneumoniae* infection for the 7-day observation period. A single injection of aureomycin was more effective than penicillin G when mice were infected with *Pneumococcus* type I (SVI). It was less effective than penicillin G when *Streptococcus hemolyticus* beta (C203) was the infecting organism.



* ALBINO SWISS MICE INFECTED INTRAPERITONEALLY WITH 1000 LD'S OF *K. PNEUMONIAE* A, *PNEUMOCOCCUS* SVI, OR *STREPTOCOCCUS* C203 AND TREATED IMMEDIATELY. SURVIVAL RECORDED FOR 7 DAYS. (10 MICE IN EACH GROUP)

FIGURE 1. Comparison of a single subcutaneous dose (50 mg./kg.) of aureomycin, polymyxin, streptomycin, and penicillin.*

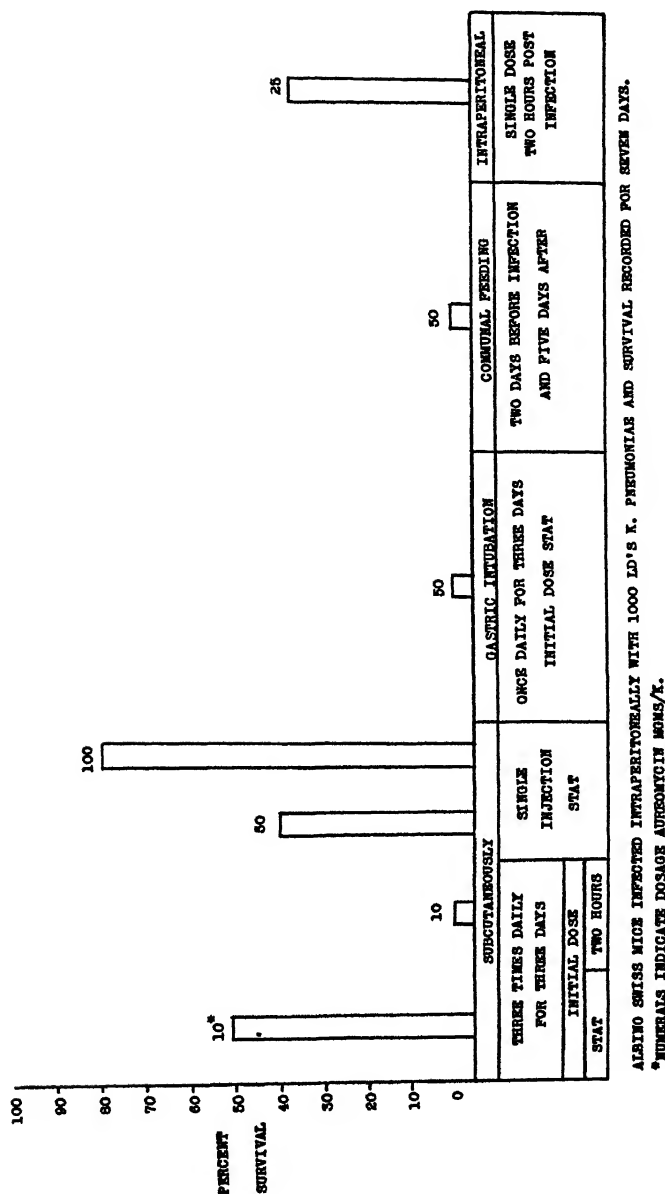


FIGURE 2. Comparison of different modes of administration of aureomycin. *K. pneumoniae* type A infection

Modes of Administration. The protection of mice infected intraperitoneally with 1000 lethal doses of *K. pneumoniae* A, *D. pneumoniae* type I (SVI), and *Streptococcus hemolyticus* beta (C203), when treated with aureomycin administered by various routes, was investigated. Com-

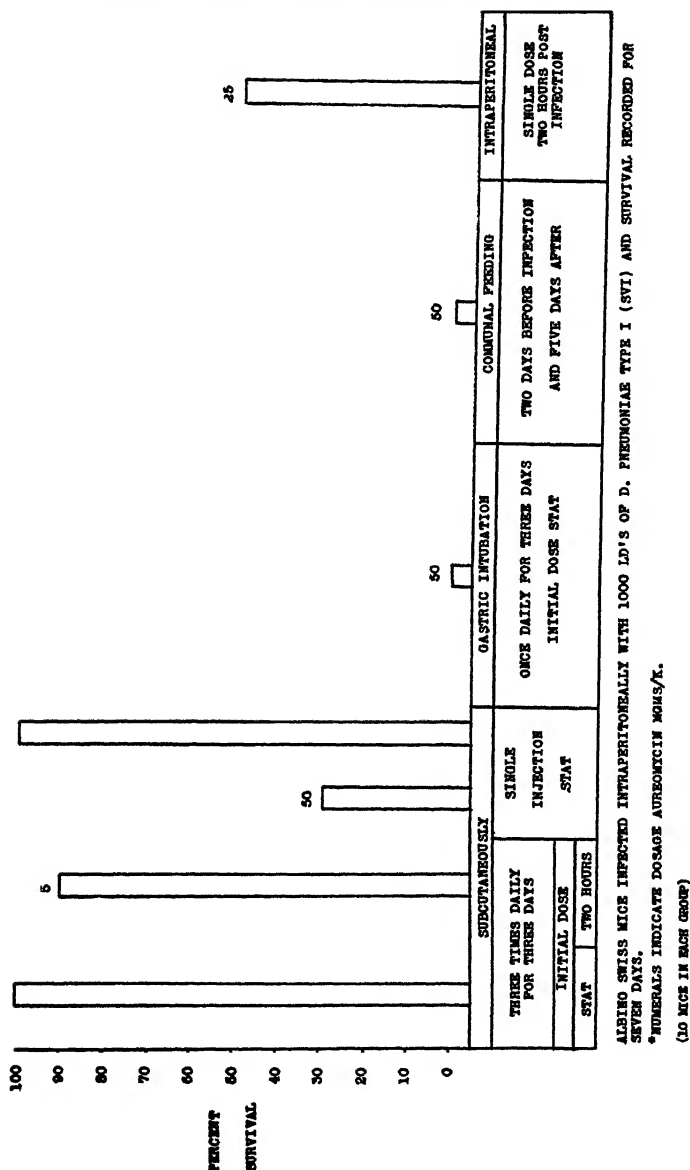


FIGURE 3. Comparison of different modes of administration of aureomycin, *D. pn. pneumoniae* type I (SVI) infection

munal feeding was begun 2 days before infection and continued for 5 days thereafter, so that a daily dose of approximately 50 mg./kg. was ingested. The only survivors observed at the end of 7 days were those mice infected with *Streptococcus hemolyticus* beta (C203). Seventy per cent protection was noted. Similarly, gastric intubation with 50 mg./kg. begun immediately after infection and continued once each day for 3 days

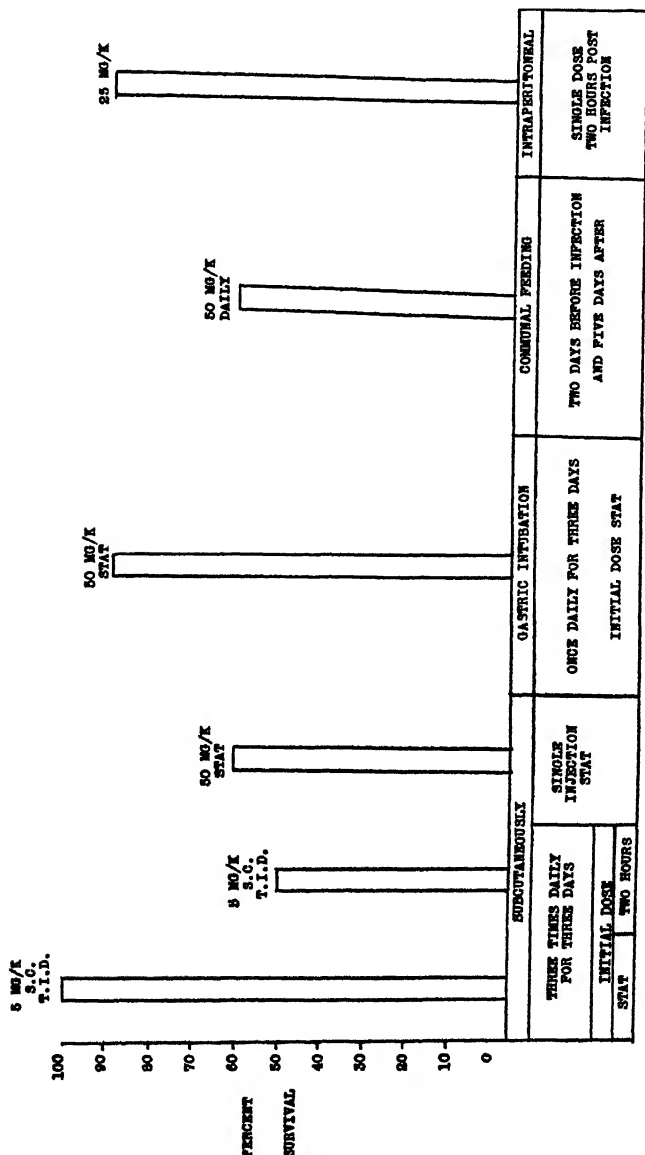


FIGURE 4. Comparison of different modes of administration of aureomycin, *Streptococcus hemolyticus* B group A (C'203) infection.

protected 90 per cent of mice infected with *Streptococcus hemolyticus* only. Subcutaneous administration of aureomycin gave some protection against all organisms both when a single or multiple dose schedule was employed. Intraperitoneal injection was more effective than the subcutaneous route, although the former therapy was not instituted until 2 hours after infection.

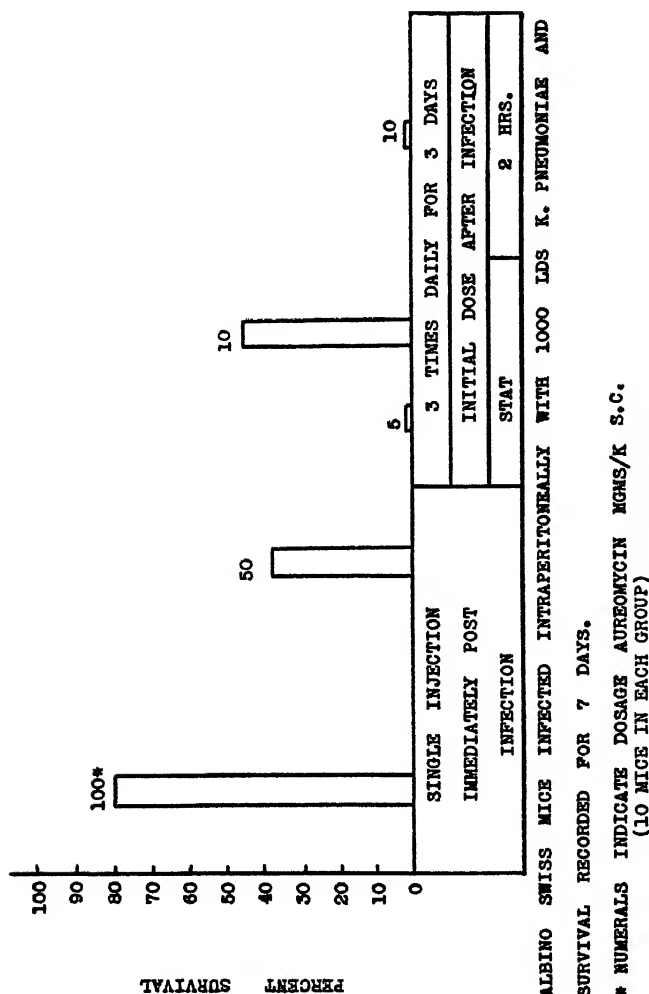


FIGURE 5. Comparison of single and multiple dose schedules in mice treated subcutaneously with aureomycin *K. pneumoniae* infection).

Single and Multiple Doses. Comparison of different dosage schedules administered subcutaneously to mice infected with 1000 LD's of *K. pneumoniae* A, *D. pneumoniae* type I, and *Streptococcus hemolyticus* beta, revealed that 5 to 10 mg./kg. given 3 times daily for 3 days resulted in more survivors than 50 mg./kg. given as a single dose. In the *Pneumococcus* infection, 5 mg./kg. 3 times a day for 3 days protected as well as a single injection of 100 mg./kg. A 2 hour delay in treatment following infection resulted in a reduction of survivors.

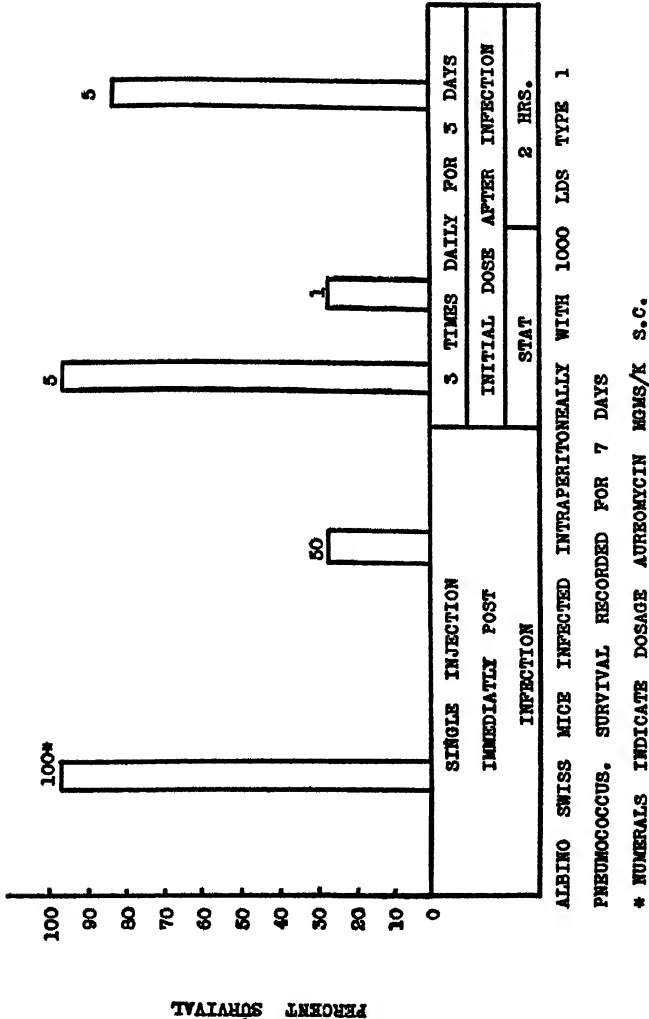


FIGURE 6. Comparison of single and multiple dose schedule in mice treated subcutaneously with aureomycin, type 1 *Pneumococcus* infection.

Comparison of Aureomycin, Penicillin, Polymyxin, and Streptomycin. Parallel experiments were performed in mice infected with the organisms mentioned above. Penicillin G was employed for the infections with the gram positive organisms, while streptomycin and polymyxin were used for the *K. pneumoniae* type A infection. Aureomycin, when compared on

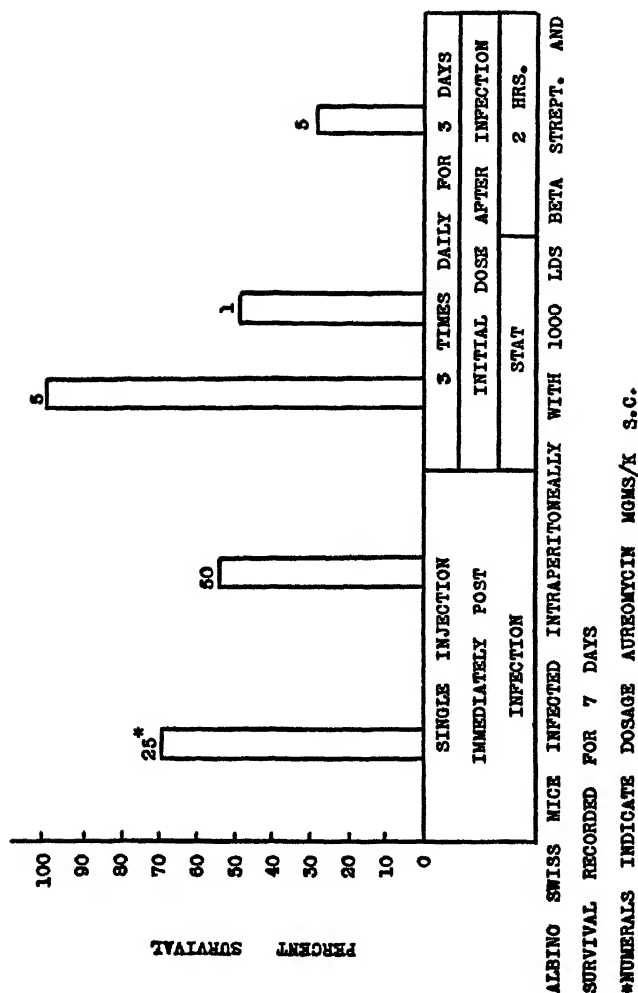


FIGURE 7. Comparison of single and multiple dose schedules in mice treated subcutaneously with aureomycin beta hemolytic streptococcus infection.

a mg./kg. basis, was as effective as penicillin when *Pneumococcus* type I was the infecting organism. Protection achieved against *Streptococcus hemolyticus* infection was somewhat inferior to penicillin. When all routes of administration were compared, the protection of mice infected with *K. pneumoniae* A and treated with aureomycin was definitely below that observed with polymyxin and streptomycin.

TABLE 5
COMPARISON OF AUREOMYCIN, POLYMYXIN AND STREPTOMYCIN
(*K. Pneumoniae* INFECTION)*

Mode of administration	Dosage schedules				Per cent survival 7 days post infection			
	Hours after infection	Injections per day	Daily dose mg./kg.	Total days treated	Aureo-mycin	Poly-myxin	Strepto-mycin	Controls
Subcutaneous	0	1	100.0	1	80			0
			5.0			60		0
			50.0		40	90	90	0
		3	15.0	3	0	100		0
			30.0		50			0
			1.5			60	20	0
	2	2	30.0	3	0			
			3.0			40	40	0
Intraperitoneal	2	1	25.0	1	40	60	80	0
Gastric intubation	0	1	50.0	3	0	70	50	0
Communal feeding	2 Days before infection	Fed continuously	50 circa	7	0	0	0	0

* Albino Swiss mice infected intraperitoneally with 1000 LD's of *K. pneumoniae* A (10 mice in each group).

TABLE 6
COMPARISON OF AUREOMYCIN AND PENICILLIN (*Pneumococcus* SVI INFECTION)*

Mode of administration	Dosage schedule				Per cent survival 7 days post infection		
	Hours after infection	Injections per day	Daily dose mg./kg.	Total days treated	Aureo-mycin	Peni-cillin	Controls
Subcutaneous	0	1	100.0	1	100	20	0
			50.0		30	0	0
		3	15.0	3	100	100	0
			3.0		30	90	0
	2	3	15.0	3	90	100	0
Intraperitoneal	2	1	25.0	1	50	0	0
Gastric intubation	0	1	50.0	3	0	0	0
Communal feeding	2 Days before infection	Fed continuously	50 circa	7	0	0	0

* 10 Albino Swiss mice infected intraperitoneally with 1000 LD's of Type I *Pneumococcus* (SVI) in each group.

TABLE 7

COMPARISON OF AUREOMYCIN AND PENICILLIN (*Streptococcus* C302 INFECTION)*

Mode of adminis- tration	Hours after infection	Dosage schedule			Per cent survival 7 days post infection		
		Injections per day	Daily dose mg./kg.	Total days treated	Aureo- mycin	Peni- cillin	Con- trols
Subcutaneous	0	1	25.0 50.0	1	70 80	90 100	10 0
		3	15.0 3.0	3	100 30	100 90	0 0
		2	3	15.0	3	50	80
Intraperi- toneal	2	1	25.0	1	90	90	0
Gastric intu- bation	0	1	50.0	3	90	60	0
Communal feeding	2 Days before infection	Fed contin- uously	50 circa	7	70	90	0

* Albino Swiss mice infected intraperitoneally with 1000 LD's of group A beta hemolytic streptococcus C203 (10 mice in each group).

Summary

1. The approximate LD₅₀ for aureomycin administered intravenously to mice is between 50 and 100 mg./kg. The LD₅₀ for subcutaneous injection is between 3000 and 4000 mg./kg. Rapid intravenous injection of 150 mg./kg. resulted in hemoglobinuria and death of a dog.

2. Repeated subcutaneous and intramuscular doses of aureomycin were well tolerated in rats and dogs except for local irritation, anorexia and weight loss. Autopsies revealed no gross or microscopic abnormalities of the viscera.

3. A rabbit tolerated 1.0 per cent aureomycin borate locally in the eye.

4. Serum levels of 0.3 to 1.25 micrograms/ml. were obtained within 1 hour after parenteral administration. No antibiotic was detected in the spinal fluid. Urine contained a high concentration of aureomycin.

5. Aureomycin administered to mice orally did not protect against *Pneumococcus* type I or *K. pneumoniae* A infection, but did protect against *Streptococcus hemolyticus* beta. Parenteral doses gave some protection against all three infections. Multiple injections resulted in more survivors than the same amount given in a single dose. Delay in treatment resulted in a reduction of survivors. Penicillin afforded slightly more protection against infection with the gram positive cocci. Polymycin and streptomycin were superior in the *K. pneumoniae* infections.

THE PHARMACOLOGY AND CLINICAL TRIAL OF AUREOMYCIN: A PRELIMINARY REPORT*†

By EMANUEL B. SCHOENBACH, MORTON S. BRYER, AND PERRIN H. LONG

*Department of Preventive Medicine,
Johns Hopkins University School of Medicine.
Baltimore, Maryland*

AUREOMYCIN (A-377) is an antibiotic derived from a strain of *Streptomyces*. *In vitro* it has been shown to possess bacteriostatic and bactericidal activity *versus* various gram positive and gram negative bacteria. Protection of laboratory animals experimentally infected with a variety of bacterial and rickettsial agents as well as the virus of lymphopathia venereum has been demonstrated. The toxicity of this antibiotic in mice and dogs was relatively low and clinical trial in selected infections was undertaken.

The drug was obtained from the Lederle Laboratories Division of the American Cyanamid Company. It was supplied as the crystalline hydrochloride salt in sterile vacuum dried ampoules each containing 20 milligrams of drug (Lot numbers 7-7845, 7-8071A and 7-8071B), and more recently in gelatin capsules each containing 50 milligrams (Lot number 7-8378). The antibiotic is a yellow powder soluble in distilled water and 5 per cent glucose in distilled water. Solutions in normal saline precipitate when a concentration greater than 1 per cent of the drug is attained. The aqueous solutions are acid and when neutralized or made alkaline deterioration of activity is rapid at room temperature.

When the drug was injected intramuscularly into patients, dissolved in distilled water (also 2.5 and 5.0 per cent glucose solution), acute pain lasting several minutes followed by a dull drawing pain which persisted for approximately one-half hour was noted. The highest concentration of aureomycin attempted was a 1 per cent solution (20 milligrams in a volume of 2.0 cc.). Upon repeated dosages the local sites became erythematous and tender. When the drug was dissolved in 1 per cent novocaine, 40 milligrams could be injected in a volume of 2.0 cc. The acute pain following the injection was then not observed, but the dull drawing pain did appear. Treatment of patients *via* the intramuscular route did not appear feasible in light of the low local tolerance to the drug.

Oral dosage of aureomycin was well tolerated. A single dose of 0.5 grams of a crude preparation was associated with loss of appetite and nausea. With more purified drug in gelatin capsules, nausea and vomiting

* Aureomycin used in these experiments was supplied by the Lederle Laboratories Division, American Cyanamid Company.

† These investigations were supported by grants received from Abbott Laboratories; Eli Lilly and Company; Lederle Laboratories Division, American Cyanamid Company; Parke, Davis and Company; and the Upjohn Company.

have been noted in only one patient on repeated occasions. The nausea in this patient was maximal following the early morning doses and was completely relieved when $\frac{1}{2}$ ounce of an aluminum hydroxide preparation (amphojel) was given with each 100 milligram dose of drug. Other patients and normal human subjects have exhibited no nausea on oral dosage of aureomycin. The highest dosage schedule employed to date has been 100 milligrams every two hours for a patient who weighed 40 kilograms.

Attempts to determine blood levels of this antibiotic have been quite disappointing. One hour following an oral dose of 0.5 grams and an intramuscular injection of 40 milligrams, a level of 0.6 micrograms per milliliter of serum was obtained. When the drug had been given for 5 to 14 days, intramuscular injection of 40 milligrams was followed by serum levels of 1.2 to 2.4 micrograms per milliliter. On oral dosage alone, no detectable blood levels have been obtained. The determination is beset with many difficulties, among which may be noted the neutralizing effect of serum proteins, the deterioration of the drug in alkaline solution, and the relatively high concentration of the antibiotic necessary for minimal inhibition of growth of the test organism, *K. pneumoniae* type A. A level of 0.6 micrograms per milliliter of serum is the lowest concentration of drug that can be measured with the present procedure.* We are trying to improve these assay methods for blood. Early trials to devise a chemical method employing the characteristic ultraviolet absorption spectrum plus extraction with organic solvents have been unsuccessful.

The urinary concentration of aureomycin after intramuscular or oral administration has been quite high. A slight delay of six to twelve hours after oral administration has been noted before urinary excretion in adequate concentration has been observed. Thence, on an oral dosage schedule of 15 mg./kg. per day urine levels of 20-80 micrograms per milliliter were obtained. When the oral dosage was 30 mg./kg. per day the urinary concentration of aureomycin was 80-160 micrograms per milliliter.

Patients and normal individuals have been given aureomycin for periods of five to twenty days on a dosage schedule of 15-30 mg./kg. per day. The age range of these patients was three years to forty-five years. Aside from the nausea noted with oral dosage in two patients and the local irritation following intramuscular administration, no evidence of toxicity has been noted. There has been no evidence of urinary impairment as measured by volume, specific gravity, albuminuria, microscopic examination, blood non-protein nitrogen levels or excretion of phenol-sulphonphthalein. No jaundice or change in icteric index, cephalin flocculation or prothrombin has been noted. Repeated hematological examination of the peripheral blood has not indicated any evidence of anemia, hemolytic process, leucopenia, or depression of platelets. Chemical ex-

* Assays of blood and urine for aureomycin were conducted by Dr. Caroline A. Chandler, Department of Preventive Medicine, Johns Hopkins University, School of Medicine.

aminations of the blood have not indicated any deviation in total protein, albumin-globulin content or ratio, cholesterol, carbon-dioxide, or alkaline phosphatase values. No idiosyncratic type of response such as rash, hives, drug fever, vertigo, vasomotor reactions, or polyneuritic phenomena have been encountered to date in this comparatively small series.

Some of the cases treated may serve to illustrate our clinical experience.

The first patient (FIGURE 1) is a 15-year-old girl who has suffered from a chronic pyelonephritis since the age of six. Therapy with sulfonamide drugs, penicillin and streptomycin had been employed singly and in combination with temporary clinical response, but since September, 1947, a continued low grade fever and pyuria were present. Repeated exacerbations of fever up to 106°, and chills had occurred. Pleurisy with effusion, jaundice, and azotemia had occurred in association with these febrile episodes. A biopsy in September 1947 of muscle and artery showed no evidence of vascular collagen disease and an exploratory abdominal laparotomy in February 1948 revealed no abnormality except for an enlarged right kidney. As all sulfonamide and antibiotic therapy had been associated with incomplete and only temporary amelioration of symptoms, she was transferred early in May, 1948, from a hospital in Florida to the Johns Hopkins Hospital for further investigation and therapy.

Examination of the urinary tract with retrograde pyelography revealed dilatation of both ureters below the uretero-pelvic junction. The calyces were somewhat dilated on the right side. A heavy growth of *coli-aerogenes* was obtained from both ureteral specimens. Repeated urine examinations were consistently abnormal with albuminuria, white blood cells in clumps, and microscopic pyurea. Urine cultures were persistently positive for *coli-aerogenes* and, at various times, *Streptococcus hemolyticus* β and enterococci also were recovered. Numerous examinations for acid-fast bacilli on smear or culture of the urine were negative. The coliform organism was resistant to 250 micrograms per milliliter of streptomycin and penicillin *in vitro*. The enterococcus was found to be sensitive to 0.9 micrograms of penicillin, but resistant to 250.0 micrograms of streptomycin per milliliter.

An attempt to sterilize the urine was made by giving 1.2 million units of penicillin G, and 2.4 grams of streptomycin intramuscularly daily. In addition, 6.0 grams of sulfadiazine and 10-12 grams of sulfasuxidine were given orally each day. While on this regime, her temperature fell to 99°-100° F., but pyuria and positive urine cultures persisted. This multiple drug treatment was continued for eighteen days. Her temperature rose to 104.6° F. twenty-four hours after the antibiotic and sulfadiazine were discontinued.

Trial of aureomycin was then considered. This antibiotic is excreted in the urine and high concentrations can be obtained. The *coli-aerogenes* organism was inhibited by 5.0 micrograms per milliliter *in vitro*. Oral aureomycin in a daily dose of 600 milligrams divided into six doses was

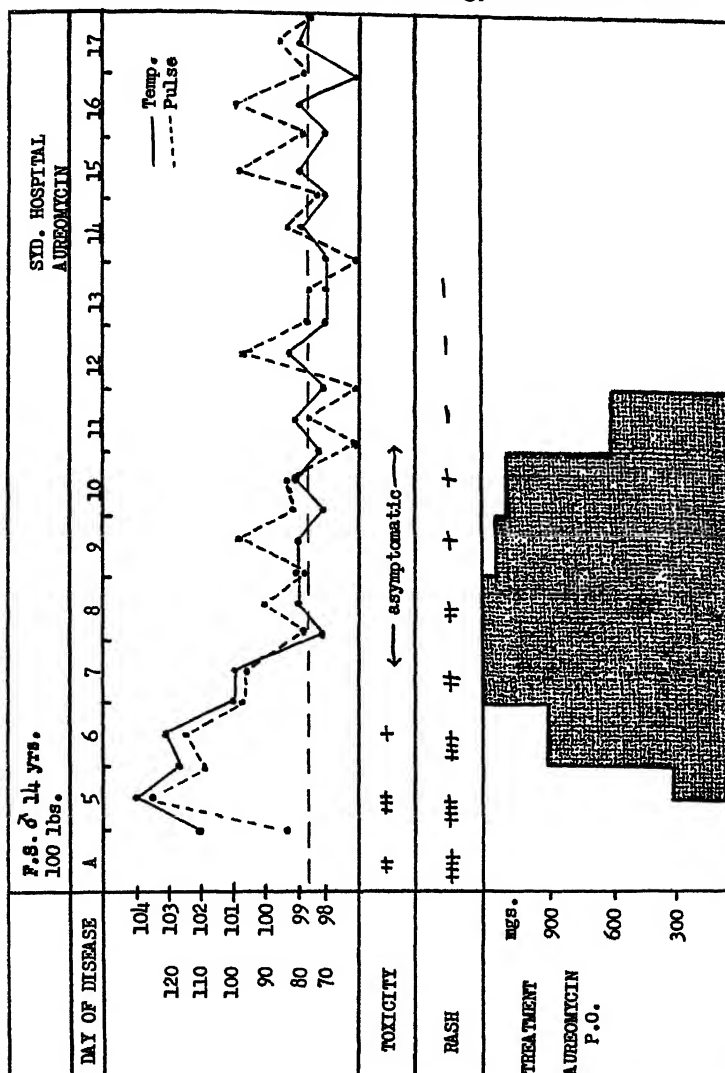


FIGURE 2. Spotted fever

became rapid, the white blood count rose to 19,000 cells and low grade fever to 101° appeared. Costo-vertebral angle tenderness, urinary frequency and urgency returned. The urine became alkaline and pyuria reappeared.

Aureomycin was discontinued. The *proteus* infection of the urine persists to date. It is felt that this patient achieved sterilization of her urinary tract infection and symptomatic relief with aureomycin, but that secondary infection with *proteus*, not susceptible to aureomycin, has supervened.

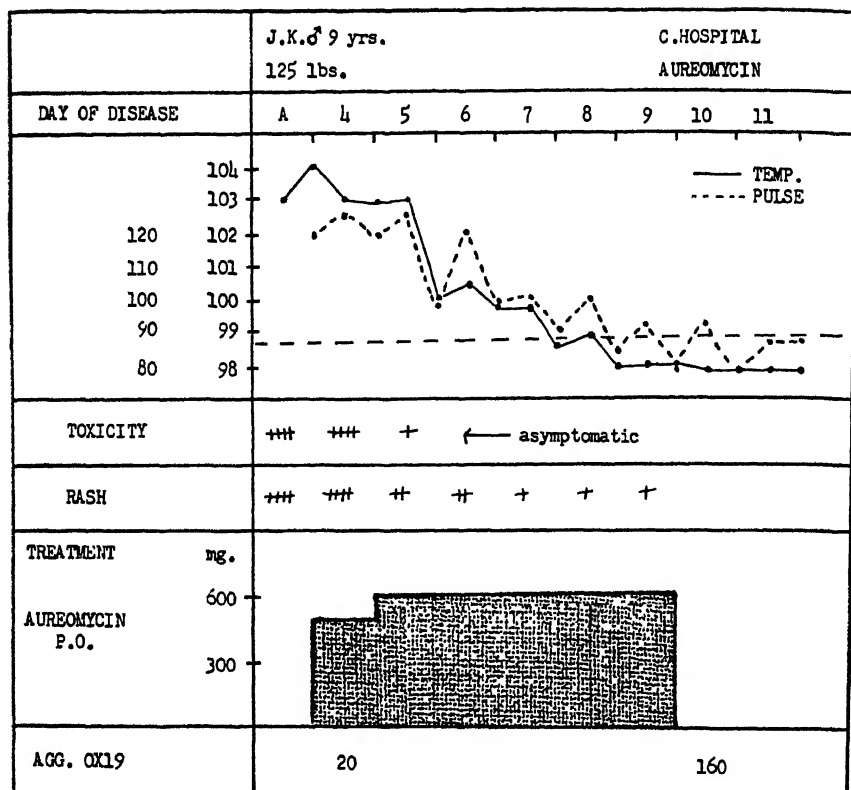


FIGURE 3. Spotted fever

This patient experienced nausea and occasional emesis while on oral aureomycin therapy. These symptoms were markedly alleviated when amphotojel was given with each dose of drug. Urine concentrations of 40-80 micrograms of aureomycin per milliliter were present on the above dosage schedule.

Another patient on the same ward with *coli-aerogenes* pyelonephritis was treated with oral aureomycin. Urinary cultures were sterile and microscopic examination revealed no abnormality within 48 hours after therapy was begun. The drug was discontinued after seven days. A *proteus* infection supervened in this patient two days after the aureomycin was discontinued.

Three cases of Rocky Mountain spotted fever, Eastern type, have been treated with oral aureomycin. Their course is outlined in FIGURES 2, 3 and 4.

A fourteen-year-old boy was admitted to the Sydenham Hospital, Baltimore, Maryland, under the care of Dr. Horace Hodes, Director,

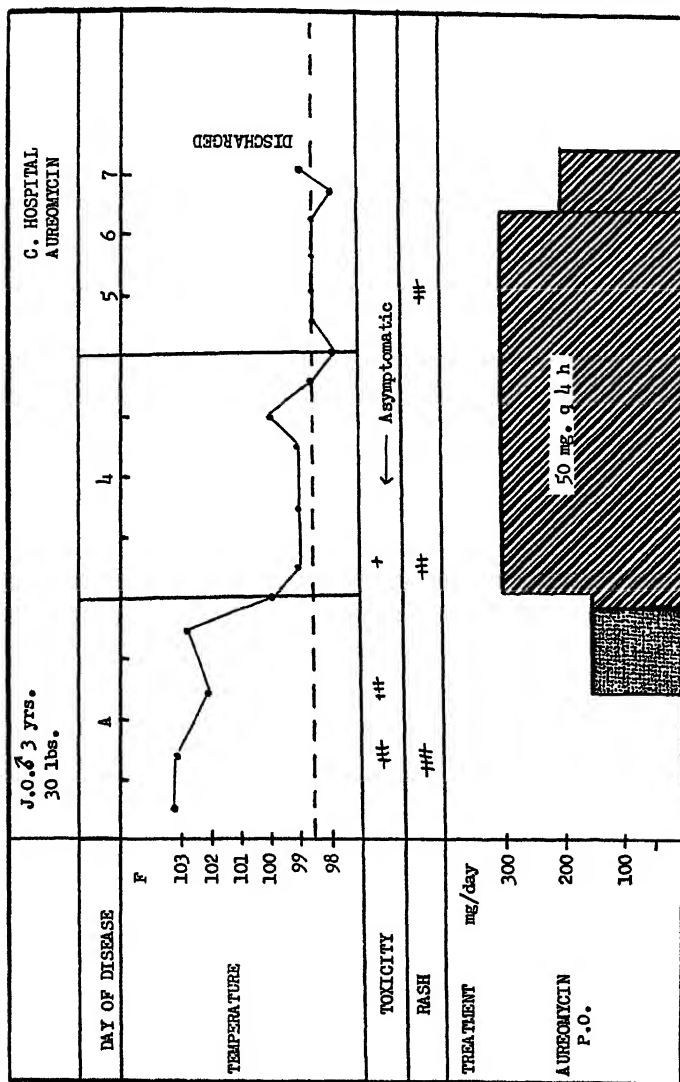


FIGURE 4. Spotted fever.

and Dr. David Karzan. He had been in Virginia during the preceding two months and had been bitten by ticks on numerous occasions. Three days before admission headache, malaise, anorexia, and high fever were noted. Penicillin 300,000 units, sulfadiazine and antipyretic capsules were prescribed. One day before admission, temperature rose to 104° F. and a rash was noted on face and ankles.

Physical examination on admission revealed an acutely ill boy. An erythematous maculopapular eruption with an occasional petechial

center was present on ankles, soles, hands, palms, and trunk. A few discrete lesions were present on the face. These lesions blanched somewhat on pressure. Nuchal rigidity was present associated with Kernig and Brudzinski signs.

There was a slight generalized lymphadenopathy and the spleen was palpable two fingerbreadths below the costal margin.

Laboratory data were:

Hemoglobin 15.5 grams per 100 milliliter

Total leucocyte count 2900

Differential leucocyte count

Polymorphonuclear cells 85%

Juvenile 11

Segmented 74

Lymphocytes 15%

Electrocardiogram—essentially normal

Spinal fluid examination—no abnormality: clear; cells 0; globulin negative.

Urine was clear, amber; specific gravity 1.028 acid and negative tests for albumin, sugar and acetone. The microscopic examination of the sediment was normal.

Blood cultures were negative.

Agglutinations *versus* *B. proteus* were positive on the fifth day of disease in a dilution of 1/40 or OX₁ and 1/80 for OX₁₈. These rose thirteen days later to OX₁—1/320 and OX₁₈—1/1280.

The temperature rose to 104° F., the day after admission, and oral aureomycin was begun that afternoon. After two initial doses of 100 milligrams given one hour apart, therapy with 100 milligrams every four hours was initiated. The next day, however, the dosage schedule was raised to 100 milligrams every two hours. The temperature fell progressively after drug therapy was begun and was normal within thirty-six hours. Convalescence was uneventful. The drug was stopped on the seventh day of treatment and the patient was discharged one week later. The splenomegaly had disappeared and he felt well.

Blood levels were not measurable in this patient, although the urinary concentration of aureomycin varied between 80 and 320 micrograms per milliliter. No nausea or other untoward symptoms were noted which might be attributed to drug therapy.

Two patients with Rocky Mountain spotted fever, Eastern type, were hospitalized at the Children's Hospital in Washington, D.C. We are indebted to Dr. Sidney Ross, Dr. C. Rice, and Dr. F. Burke of the Antibiotic Study Committee, and the staff at that hospital for their collaboration.

A nine-year-old boy, who had been intimately exposed to ticks in Virginia, became ill two days before admission with fever of 103° F. and abdominal pain. One day later a rash was noted. The child became drowsy and its temperature was maintained at 103° F. During the preceding ten days, a tick had been removed from his body and one from his clothing.

Physical examination revealed a moderately ill, obese boy with a maculopapular rash on abdomen and extremities including the palms and soles. The face was not involved. No lesions were present on the mucous

membranes of the mouth or pharynx. The spleen was not palpable and there was no lymphadenopathy. The white blood cell count was 6900 with 63 per cent polymorphonuclear leucocytes. Urine examination was normal, except for a slight albuminuria. Agglutinations were negative, except for the Weil-Felix. Agglutination of *proteus* OX₁₉ was positive in a dilution of 1/20 on admission. The latter rose to 1/160 seven days later. The rickettsial complement fixation test performed at the National Institute of Health was reported negative on June 26 and positive in a dilution of 1/64 on July 2, 1948.

Treatment with three doses of 50 milligrams of aureomycin at one hour intervals and 50 milligrams every four hours for three doses was instituted on the third day of his disease. This was then changed to 100 milligrams orally every four hours. The temperature fell by lysis and became normal on the third day of treatment. His general condition improved rapidly and twelve hours after treatment was begun he appeared alert. The drug was discontinued on the sixth day and he was discharged two days later asymptomatic.

A three-year-old infant weighing 30 pounds became ill two days before admission. Drowsiness, malaise, and fever were noted. One day before admission a rash was observed on arms, legs, elbows, and knees. This rash rapidly extended centripetally to involve other parts of the body. A tick had been removed from the right lower extremity, just above the knee, one day before his illness. Many ticks were present in the region and on the farm in Haymarket, Virginia, where he lived.

Physical examination revealed an acutely ill, lethargic infant with a maculopapular rash on all extremities including palms and soles, abdomen, chest, and back. The sclerae were injected and conjunctival blood-vessels dilated. No spleen was palpable. Oral lesions were not present.

Oral aureomycin was begun on the afternoon of admission. Fifty milligrams were given every two hours for the first two doses and thence every four hours day and night for a total period of five days. The temperature fell to normal within twelve hours after therapy was begun and he was discharged as well on the fifth hospital day.

Although none of these cases was desperately ill with signs of coma, vascular collapse, hemorrhage, or edema, it was generally believed that the response to aureomycin was indeed favorable and prompt.

A forty-five-year-old employee of a local meat packing establishment was treated with aureomycin for a persistent infection with *Brucella suis*. In January of 1948, he noted chills and fever which were alleviated somewhat by sulfadiazine. Afternoon chills and evening sweats continued, however, and he was admitted to the Sinai Hospital in Baltimore on February 14, 1948, to the service of Dr. Milton Sherry. Blood cultures were repeatedly positive for *Brucella suis*. Treatment with six grams of sulfadiazine and three grams of streptomycin daily for a period of twelve days was without effect on the clinical course or bacteremia. Polymyxin D was then given for a period of ten days at a dosage of 7½ milligrams

per kilogram of body weight per day intramuscularly divided into three doses. Blood cultures became negative promptly and remained negative for a period of four weeks. His general condition improved materially, although a low grade temperature to 100° F. and rapid blood sedimentation rate persisted. He was discharged with weekly follow-up blood cultures and physical examinations recommended.

He was readmitted to the hospital on May 3, 1948, because chills and fever to 104° F. had been noted during the preceding week and headaches, sweating, and general malaise had recurred. A blood culture drawn on April 26, 1948, was reported positive for *Brucella* four days later, although those on April 1, 12, and 19 were sterile.

Physical examination on admission revealed an acutely and chronically ill male. His face was flushed and he was sweating profusely. His liver was enlarged two fingerbreadths below the costal margin; fingers were clubbed. A soft systolic murmur at the cardiac apex was noted by some observers. Petechial and embolic lesions were present on one toe and several fingers. In the region of the right macula, there was a white area interpreted by the ophthalmological consultant as the result of a previous hemorrhage, probably of embolic nature.

Dr. Ronald M. Wood of the laboratories of the Wilmer Ophthalmological Institute, the Johns Hopkins Hospital, has performed all the *Brucella* cultures in this case. He has also collaborated with this department on *in vitro* and experimental animal studies with *Brucella*. *In vitro* tests have shown that growth of five strains of *Brucella*, *abortus* and *suis* types, was completely inhibited by 0.75 micrograms per milliliter of aureomycin. These results were obtained when the final reading was made after seventy-two hours of incubation.

Trial of aureomycin in this case was believed warranted because of the septic course, the persistently positive blood cultures, and failure of sulfadiazine, streptomycin, and polymyxin on previous occasions. After three more positive blood cultures for *Brucella suis* had been obtained, aureomycin therapy was begun. An oral dose of 0.5 grams twice a day was given for four days. This was supplemented by 20 milligrams intramuscularly on the first day, 40 milligrams on the second day, and 120 milligrams divided into three equal doses for the next two days. The oral drug was then discontinued because of anorexia and nausea. It should be noted that the oral drug (lot number 204-214) was relatively crude material. The intramuscular route of administration was maintained for a total of period of twenty days in a dosage of 40.0 milligrams four times each day. The course of this patient while on aureomycin is outlined in figure 5. His temperature became normal four days after therapy was begun. Blood cultures became sterile and have remained negative to date. The last culture drawn one month after discharge from the hospital is sterile. The patient has remained afebrile and asymptomatic. His weight, which had been 140 pounds when first readmitted in May and which had fallen to 132 pounds, is 159 pounds as of July 1. The enlarged liver has regressed

and there has been no evidence of petechial or embolic episodes. The patient has returned to work.

Blood levels while on aureomycin were determined. They varied from 0.6 micrograms per milliliter of serum to 2.4 micrograms per milliliter. Urine concentration of 10.0 to 80.0 micrograms per milliliter were measured.

Nausea and anorexia while on oral drug and local tenderness and induration following the intramuscular injections were the only untoward reactions noted. The patient will be followed for evidence of relapse.

Summary

Aureomycin, a new antibiotic, has been studied in humans. It is irritating upon parenteral administration. Oral administration is well tolerated. The drug is absorbed although blood levels are difficult to determine. It is excreted rapidly in the urine and high concentrations may be attained.

Treatment of pyelonephritis has successfully eliminated the infecting organisms of the *coli-aerogenes* group. *B. proteus* infection of the urinary tract was initiated in the presence of high concentrations of aureomycin and the drug was ineffective in the treatment of this type of urinary tract infection.

Three cases of Rocky Mountain spotted fever, Eastern type, were treated on the third and fifth day of the disease with aureomycin administered orally. All cases became afebrile within 12-72 hours and convalescence was uneventful. The drug was well tolerated in daily dosages up to 30 milligrams per kilogram of body weight. A case of *Brucella suis* infection which had not responded to sulfonamide and streptomycin therapy and which had relapsed after polymyxin therapy has become asymptomatic after treatment with oral and intramuscular aureomycin.

No toxic or allergic reactions have been observed in patients or normal individuals except for transient nausea. These may only become apparent when a larger series is treated. The preliminary results with aureomycin to date have been encouraging and more extensive clinical trial is indicated.

Addendum

Since this paper was submitted, a total of 16 cases of Rocky Mountain spotted fever have been treated with oral aureomycin. Treatment was begun on the 2nd to 8th day of disease (average 4.5) and normal temperature and clinical cure was noted 36 to 72 hours after the institution of treatment (average 2.3 days). No complications or deaths occurred. Thirteen cases in this group have been confirmed serologically with Weil-Felix and/or specific complement fixation tests.

Two additional cases of acute brucellosis have been treated, with bacteriological and clinical remission within 72 hours.

Three early cases of typhoid fever treated with aureomycin in daily

dosage of 60–100 mg./kg. per day by mouth and 3.0 to 5.0 mg./kg. per day parenterally have resulted in negative blood and stool cultures within 48 to 72 hours. Clinical response has varied, however, with defervescence occurring in 24 hours, 8 days, and 11 days respectively.

Five cases of newborn infants with multiple skin and breast abscesses due to hemolytic *Staphylococcus aureus* cleared rapidly on oral aureomycin. One of these cases also had staphylococcal infection of the lung with abscess formation and positive blood culture.

Four cases of primary atypical pneumonia have responded to oral aureomycin with prompt defervescence within 36 to 48 hours. Convalescence was uneventful.

A thirteen-year-old girl with tuberculous draining sinuses of the neck for 5 years was treated with a total of 92 grams of aureomycin orally for a period of seven weeks. Drainage ceased within 10 days and nodes decreased in size. No recurrence has been noted, although drug therapy has been discontinued for a period of three weeks.

Eight additional cases of chronic urinary tract infection which had not responded to penicillin, streptomycin or sulfonamide therapy have cleared promptly with aureomycin. The organisms identified in these cases were *B. coli-aerogenes*, *B. paracolon* and *Streptococcus fecalis*.

One case of epidemic typhus (Brill's disease) treated on the 6th day of his disease became afebrile and asymptomatic within 48 hours.

AUREOMYCIN IN OCULAR INFECTIONS

A Study of Its Spectrum

By ALSON E. BRALEY AND MURRAY SANDERS

*Department of Ophthalmology, College of Physicians and Surgeons
Columbia University, New York, N. Y.*

AUREOMYCIN is an antibiotic obtained from a mold belonging to the *Streptomyces* group. It was isolated by workers at the Lederle Laboratories of the American Cyanamid Company.¹ This antibiotic was found by these workers to be effective against both gram positive and gram negative organisms, and their work suggested a wide range of activity. It has been made available to us in two forms, Aureomycin hydrochloride and Aureomycin borate.

Aureomycin HCl is a dry, crude, crystalline substance, yellow in color. When in aqueous solution, it is an acid with a pH of 3, and is stable either in the dry form or in solution without refrigeration. In solution, it can be injected intramuscularly. The dry material is quite soluble in water or normal saline; 20 mg. will dissolve in 1 cc. to 1.5 cc. of normal saline. This material is moderately irritating when injected intramuscularly, but with the addition of a small amount of procaine hydrochloride there is little or no discomfort.

A borated salt of Aureomycin was also manufactured and dried into a fine, yellow, stable powder. In aqueous or physiological saline solution the pH is 7.5 to 7.8. It is very soluble in nearly all concentrations; however, in solution at room temperature the material loses its antibiotic activity within approximately 24 hours. Various concentrations of the borate were used, but it was found that a $\frac{1}{2}$ per cent solution was the one best tolerated. The activity of the $\frac{1}{2}$ per cent solution is retained for several days when it is stored at $+4^{\circ}$ C. Freshly prepared Aureomycin borate solution is a light golden-yellow color. After it has remained at room temperature for 24 hours it turns a golden-brown.

Wong and Cox² have shown in experimental infections with a wide range of microorganisms, rickettsiae, and viruses that Aureomycin was unusually potent and active. Since this was true experimentally, it was felt that this new antibiotic should be tried in many types of ocular infections. The antibiotic was used in this study much the same as penicillin might be used clinically. The borated salt was used locally in eyes for infections of the conjunctiva and cornea. When infection was present in the deeper structures of the eye, such as uveitis and in ocular infections associated with general disease, the local application was augmented by intramuscular injections of Aureomycin HCl.

In a previous report, 100 patients with conjunctival and corneal infections have been reported.³ It was found from this study that local use

of Aureomycin borate effectively inhibited staphylococci, pneumococci, and *H. influenzae*. It was also active against the virus of inclusion conjunctivitis, and appeared to have some value in herpes simplex corneae and to a lesser extent in epidemic keratoconjunctivitis. Since that time, additional information has been obtained regarding these same organisms as well as other ocular infections. The present study was made in an attempt to find the number of diseases and microorganisms in which

TABLE 1

RESULTS OBTAINED IN AUREOMYCIN TREATMENT OF 200 OCULAR INFECTIONS

<i>Infection</i>	<i>No. of cases</i>	<i>Clinical cure</i>	<i>No improvement</i>
<i>Conjunctivitis</i>			
<i>Staphylococcus aureus</i>			
mild	22	21	1
severe	74	73	1
<i>D. pneumoniae</i>	5	5	
<i>H. influenzae</i>	4	4	
<i>Moraxella lacunata</i> (diplobacillus of Morax-Avenfeld)	5	4	1
<i>E. coli</i>	1	1	
Follicular (etiology unknown)	14	14	
Inclusion conjunctivitis	5	5	
Trachoma	1	1	
Vernal	6	2	4
Epidemic keratoconjunctivitis	27	8	19
Molluscum contagiosum	1		1
Parinaud's conjunctivitis (leptotrichosis)	1		1
<i>Keratitis</i>			
Dendritic (herpes simplex)	6	5	1
Unclassified (etiology unknown, probably infectious)	7	3	4
Acne rosacea	3	1	2
Superficial punctate (virus?)	2	1	1
Neurotropic	2		2
Marginal, severe	1	1	
<i>Pinguecula</i>	1		1
<i>Episcleritis</i>	1		1
<i>Uveitis</i>			
Idiopathic	5	4	1
Lymphogranuloma	1	1	
Serafuloderm with uveitis and keratitis	2	2	
Sympathetic ophthalmia	2		2
Endophthalmitis, metastatic	1	1	

Aureomycin might be effective. The use of Aureomycin in 200 unselected cases showed that the antibiotic was effective in the treatment of 158 and of no value in 42. This can be further separated into 189 cases in which the infection was primarily of the conjunctiva and cornea, and 11 cases in which the infection was primarily of the uvea (see TABLE 1).

Staphylococcal Infections. Ninety-six patients with various types of staphylococcal infections were treated. The majority had blepharitis combined with a conjunctivitis and keratitis. There were, however, a fair

number of cases with marginal infiltrates and marginal ulcers associated with the chronic conjunctivitis. A few patients had horeola and chalazia complicating the blepharitis. For statistical purposes, we divided staphylococcal infections into mild and severe cases. There were 22 patients in the group considered mild, those with a blepharitis squamosa and a chronic conjunctivitis associated with a slight discharge. Severe staphylococcal infections were those with a blepharitis with an acute or sub-acute conjunctivitis combined with corneal changes such as superficial punctate keratitis, marginal infiltrates, or marginal ulcers. The evaluation of the results of Aureomycin in staphylococcal infections was difficult because of the tendency for recurrence of all types of staphylococcal infections and the inclination of the patient to discontinue treatment as soon as the acute symptoms subsided. In order to evaluate the effectiveness of the antibiotic, an arbitrary 48 hours was chosen as the time limit for disappearance of symptoms. If there was no improvement in the objective appearance or symptoms, the action of the antibiotic was considered unfavorable. There is little doubt, as reported in the previous publication, that Aureomycin is very active against the staphylococcus, whether the disease produced by it is mild or severe. From experience with these patients, it seems to be as potent as local penicillin (1,000 units per cc). Staphylococcal conjunctivitis had a tendency to recur when therapy was discontinued. Recurrences were about the same with Aureomycin as with other local antibiotics. Aureomycin was potent in patients with staphylococcal conjunctivitis who had developed a sensitivity to the local use of penicillin.

Staphylococcal infections require local therapy for a long period of time and since a transition occurs in Aureomycin borate solution, an attempt was made to prepare an ointment with both the hydrochloride and the borate. The hydrochloride ointment was irritating to nearly all of the patients, although it seemed to be active against the infection. The ointment prepared with the borate was non-irritating, but it probably lost its activity after a short period. Further investigation will be necessary to determine the best means of retaining this antibiotic effect for the required prolonged use in staphylococcal infections.

There were only two unfavorable responses to Aureomycin in the entire group. The organisms could have been assayed to determine the amount of the antibiotic necessary to inactivate them, but it is possible that the patients did not use the antibiotic as directed.

Pneumococcal Conjunctivitis. The second largest group of patients with bacterial conjunctivitis was infected by the pneumococcus. Only 5 patients with pneumococcal conjunctivitis have been treated; there were excellent results and no recurrences.

Influenzal Conjunctivitis. The response of 4 patients from whom influenza bacilli were cultured from the conjunctiva was rapid and efficient. In all of the patients, the purulent discharge was entirely gone within

24 hours. It was difficult to evaluate the antibody response to influenzal conjunctivitis, since the duration of this infection without treatment varied considerably. From the appearance of the conjunctiva, however, we felt that Aureomycin was potent against *H. influenzae*. Further investigation will be necessary, particularly in influenzal meningitis, to determine the degree of potency.

Diplobacillary Infections. The diplobacillus of Morax-Axenfeld has been an organism which has not responded well to any form of therapy. In 4 cases the symptoms were relieved and no bacteria found on scrapings of the conjunctiva after local use of Aureomycin. In one patient, a recurrence developed in one eye shortly after the antibiotic was discontinued. From all clinical appearances, however, Aureomycin is much more effective against the diplobacillus than any other therapy.

E. coli. Aureomycin borate was used in only one case of infection of the conjunctiva by *E. coli*. A pure culture of *E. coli* was obtained from the conjunctiva after several months of conjunctivitis. Following the use of this antibiotic for 24 hours, conjunctival symptoms were entirely gone and the conjunctiva was free of organisms.

Viral Infections. Two viruses involving the conjunctiva and cornea have been treated with Aureomycin, 6 cases of dendritic keratitis, and 27 cases of epidemic keratoconjunctivitis. Clinically, Aureomycin appears to be more useful in the herpes corneae virus than in epidemic keratoconjunctivitis.

Dendritic Keratitis. All 6 of the patients with dendritic keratitis had an associated beginning involvement of the corneal stroma with an area of infiltration under the dendritic figure. In 5 of these patients, the results after use of Aureomycin were rapid and beneficial. The ulcer was healed in 24 hours and there was no increase in the size of the infiltrate beneath the ulcer, and in most of the patients the infiltrate disappeared entirely. In one patient who had a recurrent dendritic keratitis with a large disciform keratitis there was no appreciable change in the involvement of the corneal stroma after Aureomycin. From our experience thus far with the antibiotic, it seems to be a valuable therapeutic agent for dendritic keratitis.

Epidemic Keratoconjunctivitis. The use of Aureomycin in the treatment of epidemic keratoconjunctivitis does not appear effective from the data presented. Twenty-seven patients with the infection have been followed after the use of the antibiotic. There were many more patients with epidemic keratoconjunctivitis who had used Aureomycin, but their follow-up was not entirely satisfactory and they cannot be included in the series. The course of the disease was not affected in 19 of the 27 patients, but 8 showed a definitely favorable reaction. These patients taught us a good deal as to how this new antibiotic should be used. Certainly in epidemic keratoconjunctivitis, Aureomycin must be used before

corneal opacities begin, since in several of the 19 patients it was started after corneal opacities were first noted. It appears that if Aureomycin can be started before the third to fifth day of the disease and used continually for at least a week or ten days there is some beneficial effect. If the material is instilled in the conjunctival sac every one-half to one hour and continued, even though the symptoms and edema of the conjunctiva increase, there will certainly be some favorable results. Most of the 8 favorable cases developed one or two typical corneal opacities at about the same time as they would had the disease been allowed to run its normal course, but the conjunctival findings and the corneal changes were considerably decreased as compared to control cases. Even though Aureomycin was used, a few developed conjunctival scars and minimal symblepharon. Patients who used the antibiotic a few times and discontinued its use after 24 or 48 hours because of its slight irritability received no benefit from the antibiotic. The method of choice, therefore, in treatment of epidemic keratoconjunctivitis should be the instillation of Aureomycin borate every hour for at least ten days with a fresh supply of the antibiotic given at 48-hour intervals. Blood has been obtained from the patients for neutralization of the virus to determine the presence of antibodies.

Follicular Conjunctivitis, Unknown Etiology. The evaluation of the use of Aureomycin in follicular conjunctivitis is rather difficult, since this group contains a number of patients with what clinically appeared to be a Beal's type of conjunctivitis. This type of follicular conjunctivitis is notoriously inconsistent in the duration of symptoms. With Aureomycin, however, none of the cases of follicular conjunctivitis lasted more than 48 hours.

Inclusion Conjunctivitis and Trachoma. Local Aureomycin was used in 5 cases of inclusion conjunctivitis. Most of these were in the new-born infant after the disease had been present for from five days to two weeks. The response to Aureomycin was prompt and the purulent discharge was entirely gone after 24 hours of its use. The conjunctiva returned to normal within three days to one week. Some of the cases have been followed with daily scrapings from the conjunctiva; no inclusion bodies could be found after 24 hours on Aureomycin.

One case of trachoma III which had been recurrent for twenty years was treated. This patient had developed anuria on sulfanilamide. Following treatment with local Aureomycin there was rapid disappearance of corneal infiltrates and conjunctival symptoms. No inclusion bodies could be demonstrated from the conjunctival scrapings. From the effect of Aureomycin on inclusion conjunctivitis and lymphogranuloma, we feel that it should also be effective in trachoma.

Molluscum Contagiosum. One patient with molluscum contagiosum has been treated with Aureomycin locally. There was no effect on the tumor or on the appearance of the conjunctiva.

Keratitis. Fifteen cases of various types of keratitis have been treated. Some of these have been remarkable, particularly one patient with a severe marginal keratitis of unknown etiology. This patient had marked thinning of the periphery of the cornea, and the central cornea was necrotic. His vision was reduced to hand movements in one eye and 20/400 in the other eye. After local and parenteral Aureomycin the process stopped and the central portion of the cornea became clear while the periphery vascularized. The vision was restored to 20/25 in one eye and 20/20 in the other.

Three cases of acne rosacea keratitis have been treated, one with some beneficial effect and two without any effect. It is doubtful that any of the various types of keratitis profunda or severe types of keratitis are particularly benefitted by Aureomycin although 6 of the 15 cases improved under Aureomycin therapy.

Two patients with neurotropic type of keratitis were unaffected by Aureomycin.

Parinaud's Conjunctivitis. Only one patient with Parinaud's conjunctivitis was treated with Aureomycin with no effect on the progress of the disease.

Vernal Conjunctivitis. Six cases of vernal conjunctivitis have been treated with varying amounts of local Aureomycin. Four were unchanged, but two showed a disappearance of the filmy membrane and some relief of symptoms, although there was no change in the papillary hypertrophy in the conjunctiva. It is probable that the improvement was due to the antibiotic action on the secondary infection.

Uveitis. All of the patients with uveitis have been treated with local horate salt and intramuscular Aureomycin HCl. Of the 11 patients with various types of uveitis, 8 have shown beneficial results, while 3 have remained unchanged. The most striking in this group are 2 patients with scleruloderm combined with a keratitis and uveitis. Not only was the eye restored to normal vision, but there was also marked improvement in the draining sinuses and skin reaction associated with the underlying lymph node disease. In one of these patients, tubercle bacilli were demonstrated in biopsies of the lymph nodes, and in the other, the biopsy and guinea pig inoculation was unsatisfactory. There are under observation at the present time two other patients with tuberculous lymphadenopathy and keratitis. The response of the cornea has been prompt, but there has been no change in the lymphadenopathy so that these cases have not been included in this study. The response to Aureomycin in these cases is certainly as satisfactory as with streptomycin.

One patient with a recurrent unilateral uveitis in whom a positive Frei test was found, responded dramatically to treatment with Aureomycin. She developed a recurrence two weeks after receiving 100 mg. intramuscularly and was given another course of 100 mg. with complete recovery.

Five cases of uveitis of unknown etiology have received Aureomycin; 4 have shown definite improvement while one has remained unchanged. Several more patients are under observation at the present time, but it is too early to draw any conclusions as to the efficiency of the antibiotic.

Two cases of sympathetic ophthalmia were given Aureomycin without any beneficial result, although one patient insisted on receiving it for a long period of time because he stated that his eyes felt better. There was no change in the clinical appearance of the eye.

One patient with a metastatic endophthalmitis of unknown etiology has been treated. She had developed a large abscess in the vitreous hump with pus in the anterior chamber following considerable abdominal surgery. She was given 300 mg. intramuscular Aureomycin and local drops after which all of the pus disappeared. A red reflex could be obtained from the fundus, but because of the vitreous opacities it was impossible to see the fundus details at the time of her discharge from the hospital. The eye has remained quiet since that time.

Comments

We have previously reported on the antibiotic properties of local Aureomycin in staphylococcal, pneumococcal, and influenzal conjunctivitis. To this group we are adding the diplobacillus of Morax-Axenfeld and *E. coli*. Several diseases of unknown etiology were treated in an attempt to determine the spectrum of the antibiotic. The validity of Aureomycin in some of the virus infections was not anticipated. From the experiments of Wong and Cox² we learned that it should be of value in the psittacosis-lymphogranuloma group and in rickettsial diseases. In our experience, Aureomycin was excellent in the treatment of inclusion conjunctivitis. Because of the similarity between the virus of trachoma and the virus of inclusion, it should be a useful therapy for trachoma (see TABLE 2).

Aureomycin seems to have some antibiotic properties to the virus of herpes simplex. The results of its use on herpes infections of the cornea were striking. Experimental herpes simplex infections will be reported later, although at the present time they indicate that Aureomycin has some anti-viral properties.

Its use in epidemic keratoconjunctivitis is not entirely satisfactory, but there is a strong indication from several patients that if the antibiotic is used properly, considerable benefit may be expected. After Aureomycin borate had been used for a period of 48 hours, it became somewhat irritating to the conjunctiva, in cases of epidemic keratoconjunctivitis particularly, and many patients thought it produced a more severe conjunctivitis. We found that the applications of the antibiotic must be continued in spite of increased symptoms in order for it to be beneficial. There is no change in epidemic keratoconjunctivitis much short of a week. If the material was used conscientiously by the patient,

TABLE 2

OCULAR INFECTIONS WHICH SHOULD RESPOND WELL TO AUREOMYCIN THERAPY

<i>Virus infections:</i> inclusion conjunctivitis [*] trachoma [*] lymphogranuloma venereum ^{**} herpes simplex corneae [*] follicular conjunctivitis [*]	<i>Gonococcus infections.</i> conjunctivitis iridocyclitis <i>B. pyocyaneus infections</i> ulcers <i>B. proteus infections.</i> conjunctivitis <i>Coliform group infections.</i> conjunctivitis [*] <i>H. influenzae infections.</i> conjunctivitis [*] ulcers ^{**} orbital cellulitis
<i>Hemolytic streptococcus infections.</i> conjunctivitis ^{**} (membranous) corneal ulcers ^{**} endophthalmitis [†] orbital cellulitis ^{**} impetigo [*]	<i>Diplobacillus (Morax-Axenfeld)</i> conjunctivitis [*] ulcers ^{**} <i>Friedlander bacillus infections:</i> ulcers ^{**} conjunctivitis [*] meibomitis ^{**} dacryocystitis
<i>Staphylococcus infections.</i> dacryocystitis conjunctivitis [*] ulcers [†] endophthalmitis ^{**} blepharitis [*] orbital cellulitis impetigo [*]	<i>Meningococcus infections.</i> endophthalmitis conjunctivitis
<i>Pneumococcus infections.</i> dacryocystitis conjunctivitis [*] ulcers ^{**} endophthalmitis orbital cellulitis	

^{*} Local therapy preferred.^{**} Combined intramuscular and local therapy preferred.

TABLE 3

OCULAR INFECTIONS IN WHICH AUREOMYCIN MAY BE OF VALUE AND DESERVES TRIAL

<i>Virus infections:</i> epidemic keratoconjunctivitis [†] herpes zoster ^{**} herpes simplex corneae [†]	<i>Brucella melitensis-abortus-suis:</i> keratitis uveitis choroiditis
<i>Tuberculosis:</i> conjunctivitis [†] (ulcers) uveitis ^{**} keratitis serafuloderma ^{**} kerato-uvetis	<i>Moraxella duplex (diplobacillus of Petit):</i> central ulcers <i>H. Ducreyji</i> soft chancre of lid or conjunctiva <i>Syphilis:</i> chancre of lid choroiditis optic atrophy <i>Keratitis (marginal) unknown etiology</i> <i>Uveitis (idiopathic)</i>
<i>Non-hemolytic streptococcus infections:</i> orbital cellulitis endophthalmitis corneal ulcers	

^{*} Local therapy preferred.^{**} Combined intramuscular and local therapy preferred.

the period of morbidity was shortened from approximately three weeks to a period of ten days. The corneal opacities which developed were usually minimal as compared to control cases (see TABLE 3).

Aureomycin had no effect on the virus of molluscum contagiosum.

Aureomycin was surprisingly effective in 14 cases of follicular conjunctivitis. In none of these was it possible to determine the etiology.

Six cases of vernal conjunctivitis were treated and no particular result was anticipated. It is well known that any of the antibiotics will have some tendency to give a certain degree of subjective improvement, and it is doubtful if Aureomycin will have any effect on vernal conjunctivitis.

Although one case of Parinaud's conjunctivitis due to the leptothrix was treated, no beneficial effect was anticipated because of the similarity between the leptothrix and the actinomyces form which Aureomycin is made.

In the treatment of uveitis and keratitis, it was not possible to determine in advance which patients would be improved. There is little doubt in our minds that there was marked improvement in the patients with scrofulous keratitis and scrofuloderm. There is also little doubt that the antibiotic produced improvement in several cases of uveitis of unknown etiology.

TABLE 4 lists some infections in which Aureomycin is not expected to be of value.

TABLE 4

OCULAR INFECTIONS IN WHICH AUREOMYCIN PROBABLY IS OF NO VALUE

Erythema multiforme:	Sympathetic ophthalmia
conjunctivitis	Vernal conjunctivitis
keratitis	Molluscum contagiosum
Ocular pemphigus	Mooren's ulcer
Parinaud's conjunctivitis:	Streptothrix concretions
leptotrichos	

Summary

Aureomycin borate has been used locally, and Aureomycin HCl has been used intramuscularly in 200 patients with a wide range of ocular infections. The local use of $\frac{1}{2}$ per cent solution produced no damage to the conjunctiva or cornea. This antibiotic was found to be effective against some of the gram positive cocci and several gram negative bacilli. It was also found to be an effective therapeutic agent in inclusion conjunctivitis and in herpes simplex of the cornea. Its therapeutic effect in epidemic keratoconjunctivitis will require further investigation before results can be evaluated. It is, however, more effective in epidemic keratoconjunctivitis than any of the other antibiotics or drugs tried. The intramuscular administration of Aureomycin HCl did not give rise to any toxic reactions and in only one individual was any general effect noted. This patient developed a secondary anemia which was easily controlled by the administration of iron. The HCl is somewhat irritating on intra-

muscular injection, but this irritation can be controlled by the addition of a small amount of procaine hydrochloride. There is some indication that Aureomycin may be a valuable antibiotic in the treatment of uveitis.

Aureomycin has a wide spectrum of activity in ocular infections.

References

1. DUGGAR, B. M., & C. W. HESSELTINE. Personal communication, to be published.
2. WONG, SAM C., & HERALD A. COX. 1948. *Ann. N. Y. Acad. Sci.* 51(2): 290-305.
3. BRALEY, A. E., & MURRAY SANDERS. A preliminary report on Aureomycin, a new antibiotic with viricidal properties. To be published in *J. A. M. A.*

ACTION OF AUREOMYCIN AGAINST EXPERIMENTAL RICKETTSIAL AND VIRAL INFECTIONS

By SAM C. WONG AND HERALD R. COX

*Section of Viral and Rickettsial Research,
Lederle Laboratories Division, American Cyanamid Company,
Pearl River, N. Y.*

TIME does not permit for a review of the literature on the treatment of viral and rickettsial infections with chemical and antibiotic substances. The results of such investigations have been reported recently by several groups of workers.

In this preliminary report are presented some of the data obtained with the newly described antibiotic, Aureomycin, in experimental infections of embryonated eggs, mice, and guinea pigs with the agents of lymphogranuloma venereum, psittacosis, murine typhus, epidemic typhus, Rocky Mountain spotted fever, Q fever, rickettsialpox, and scrub typhus.

Materials and Methods

In most instances the antibiotic was resuspended in one-twentieth molar phosphate buffer pH 7.0. Buffered suspensions of this type, containing 5 mg. of Aureomycin per ml., have a final pH of 6.8. Any solution left over was usually discarded, although subsequent studies indicate that if the unused material is frozen immediately, the potency appears unimpaired for at least 2 weeks.

Chemotherapy Experiments in Eggs. The following method was used to determine if the chemotherapeutic agent under investigation showed antiviral or antirickettsial activity. Seven-day-old embryonated hens' eggs were inoculated into the yolk sac with 0.2 ml., containing the desired dosage of Aureomycin. The control group received the same amount of buffered saline. Thirty minutes later all groups of eggs were reinjected into the yolk sac with 0.1 ml. of the various ten-fold dilutions of infectious inoculum. All dilutions of the infectious inoculum were made with cold 0.1 per cent casein solution, pH 7.0, and all inoculum dilutions were kept refrigerated until used. All eggs were sealed with collodion, stored in an incubator at 96° F. and candled at least once a day for a total of 13 days.

In most instances, deaths of embryos before the fourth day were considered to be non-specific and were not included in the analysis of the results. All embryos dying on the fourth day and thereafter were thoroughly examined for the presence of elementary bodies or rickettsiae, using yolk-sac or amniotic-sac smears stained by the Macchiavello method. All embryos surviving after the 13th day were examined in the same manner.

Chemotherapy Experiments in Mice. Swiss albino mice, weighing from 14 to 16 grams, were infected by the intracerebral or intraperitoneal route with 0.03 ml. or 0.25 ml., respectively, of an appropriate dilution of infectious inoculum. Aureomycin was administered to mice by mouth or injected subcutaneously, usually in a single dose given once a day. Mice treated orally were given the drug by means of a 1 or 2 ml. tuberculin syringe fitted with a round-end 18 gauge needle, 1½ inches long. The mice were grasped firmly by the back of the head and turned over with the head, slightly elevated, held away from the operator. The full length of the needle was slid slowly past the esophagus into the stomach and the required amount of Aureomycin was then given.

Chemotherapy Experiments in Guinea Pigs. Guinea pigs, weighing 450 to 500 grams, were injected intraperitoneally with 1 ml. each of the appropriate dilution of infectious inoculum. All dilutions were made with 0.1 per cent casein solution pH 7.0. Rectal temperatures were taken for 5 days prior to infection and were continued for at least 2 weeks post infection. Aureomycin was given subcutaneously, as a rule once a day, and in most cases consisted of a 3 mg. dose in 1.5 ml. amounts.

Results

TABLE 1, which represents a typical experiment carried out in em-

TABLE 1

PSITTACOSIS IN EMBRYONATED EGGS

DILUTION OF INOCULUM	SURVIVAL RATIO ON 13 TH DAY		PRESENCE OF <i>M. PSITTACII</i> BY MICROSCOPIC EXAMINATION	
	TREATED*	CONTROLS	TREATED	CONTROLS
10 ⁻¹	8/8		0/8	
10 ⁻²	9/9		0/9	
10 ⁻³	9/10	0/10	0/10	10/10
10 ⁻⁴	8/9	6/10	0/9	10/10
10 ⁻⁵	8/10	0/10	0/10	10/10
10 ⁻⁶	7/9	0/9	0/9	10/10
10 ⁻⁷		0/9		10/10
10 ⁻⁸		3/9		5/9

* A SINGLE INJECTION OF 1 MG OF DRUG GIVEN 30 MINUTES PRIOR TO INOCULATION

bryonated hens' eggs, shows the results obtained with Aureomycin in psittacosis infection. A single injection of 1 mg. of Aureomycin was given

into the yolk sac 30 minutes prior to inoculation with serial ten-fold dilutions of psittacosis (6 BC strain) infected yolk sac. The results show that 44 of the 55 treated embryos survived the infection through the 13th day, whereas only 3 of 57 control embryos survived. That the protection afforded was not border-line is shown by the fact that the drug protected all embryos that received the 10^{-1} and 10^{-2} dilutions of infectious inoculum, equivalent approximately to 10 million and 1 million lethal doses, respectively. Secondly, it is seen that it was not possible to demonstrate elementary bodies in any of the drug treated embryo yolk sacs, although 55 of the 59 control embryos were positive.

The results shown here were found to hold true essentially for all the viruses of the psittacosis-lymphogranuloma group (lymphogranuloma venereum, psittacosis, SF strain human pneumonitis, mouse pneumonitis, feline pneumonitis, and meningo-pneumonitis F97 strain), as well as the rickettsiae of murine typhus, epidemic typhus, scrub typhus (Karp strain), Rocky Mountain spotted fever, boutonneuse fever, South African tick-bite fever, rickettsialpox, North Queensland tick typhus, and Q fever (original Montana Nine-Mile strain, California bovine strain, the Dyer strain, Lynn strain, Italian Henzerling strain, and two Swiss strains).

It should be pointed out that although no elementary or rickettsial bodies could be demonstrated microscopically in the yolk sacs of Aureomycin-treated eggs in any of the above named infections, when 4 or more yolk sacs from living treated-embryos that received the more concentrated inocula were pooled to make a 10 per cent suspension and injected into other embryonated eggs, elementary or rickettsial bodies could be demonstrated in most instances in the yolk sacs of the first passage eggs. Hence, it is apparent that chick embryos that received massive infecting doses were not entirely freed of the infectious agent by a single injection of Aureomycin. In all probability a quantitative relationship exists between the dosage of the infecting agent and the dosage of Aureomycin used. Further experiments along these lines are in progress.

More recent experiments with psittacosis and epidemic typhus have shown that treatment of infected embryonated eggs with Aureomycin may be delayed as long as 24 hours post infection with marked beneficial results.

Effect of Aureomycin IN VITRO. In experiments carried out with epidemic typhus, murine typhus, Q fever and psittacosis, it was found that Aureomycin showed little or no direct virucidal or rickettsiacidal activity. This was demonstrated in experiments in which 2 portions of the infectious agent under test were prepared; one contained the infectious agent suspended in 0.1 per cent casein solution pH 7.0, whereas the other contained the infectious agent in 0.1 per cent casein solution in which was dissolved 1 mg. of Aureomycin per ml. The final concentration of infected yolk-sac suspension in the mixtures varied from 1:100 to 1:1000. The

mixtures were held at room temperature for 30 minutes and then titrated in embryonated eggs. In all instances the LD₅₀ titers of the two portions were practically identical. In experiments of this type, virucidal or rickettsiacidal activity apparently was observed only in those eggs that received the original mixture in which the antibiotic was present in highest concentration

TABLE 2 shows the results obtained when mice were injected intra-

TABLE 2
PSITTACOSIS

INOCULATION OF YOLK-SAC SUSP		TREATMENT* 1 MG DAILY	SURVIVAL RATIO OF MICE 17 DAYS AFTER INOC	
ROUTE	DILUTION	NO OF DAYS	TREATED	CONTROLS
I P	10 ⁻³ **	1	4/10	
		3	10/10	
		5	9/10	
		7	8/9	
				0/10 ...

* TREATMENT STARTED 24 HOURS AFTER INOCULATION

** 10⁻³ = 47000 LD₅₀

peritoneally with approximately 47,000 LD₅₀ of psittacosis infected yolk-sac inoculum and treated with Aureomycin by the subcutaneous route with 1 mg. per day for 1, 3, 5, and 7 days, starting treatment 24 hours after infection. The results show that a minimum of 3 injections of 1 mg. each daily were required to give complete protection to the treated mice. However, a single injection of 1 mg. showed some protection since 4 out of 10 mice survived in this group, whereas all the controls died.

TABLE 3 shows the results obtained when mice were injected *intracerebrally* with graded doses of lymphogranuloma venereum virus and treated with Aureomycin by the subcutaneous route with 1 mg. per day for 7 days, starting 24 hours post infection. It is seen that 39 of the 40 mice treated were completely protected, whereas all 39 mice receiving the same infectious dilutions died. Similar results have been obtained with mice infected intracerebrally with the 6 BC strain of psittacosis virus. These findings are noteworthy in view of the fact that Smadel and Jackson (1948) reported that Chloromycetin shows no beneficial effect in mice infected intracerebrally with either of the above viruses. These results indicate that Aureomycin is more active and that it must be able to pass the blood-brain barrier in order to protect mice against intracerebral infection with the psittacosis-lymphogranuloma group of viruses.

TABLE 3

LYMPHOGRANULOMA VENEREUM

INOCULATION OF YOLK-SAC SUSP.		TREATMENT*	SURVIVAL RATIO OF MICE	
ROUTE	DILUTION	DAILY	TREATED	CONTROLS
		MG		
	10^{-2} **		9/9	0/11
	10^{-3}		10/10	0/9
I C	10^{-4}	10	11/11	0/9
	10^{-5}		9/10	0/10
	10^{-6}			4/10

* TREATMENT STARTED 24 HOURS AFTER INOCULATION
AND CONTINUED FOR 7 DAYS

** $10^{-2} = 13000 \text{ LD}_{50}$

TABLE 4 shows the results obtained when mice were injected intracerebrally with approximately 500 LD_{50} of lymphogranuloma venereum virus and treated with Aureomycin by the subcutaneous route with 1 mg per day for 3 days, starting treatment 24, 48, 72, and 96 hours post infection. It is seen that of the 10 active mice treated within 24 hours, all survived. Those treated at 48 hours, at which time 2 were sick and 6 were active, all survived. With those treated at 72 hours, at which time all 9 were sick, 7 of the 9 survived. Those treated at 96 hours, at which time all 8 had been sick for 2 days, only 1 of the 8 survived. The results obtained in the last group bring up the question—would more mice have been saved had the daily dosage been increased or had the number of doses per day been increased?

TABLE 4

LYMPHOGRANULOMA VENEREUM

INOCULATION OF YOLK-SAC SUSP.		TREATMENT*	NO. OF MICE IN GROUP		SURVIVAL RATIO
ROUTE	DILUTION	STARTED	AT START	TIME OF TREATMENT	
I C	10^{-2} **	24 HOURS	10	10 ACTIVE	10/10
		48 "	8	2 SICK 6 ACTIVE	8/8
		72 "	9	9 SICK	7/9
		96 "	8	8 SICK (2 DAYS)	1/8
		CONTROLS	10		0/10

* 1 MG DAILY FOR 3 DAYS

** 10^{-2} DILUTION = 513 LD_{50}

TABLE 5 shows the results obtained when mice were injected intracerebrally with graded doses of lymphogranuloma venereum virus and treated with Aureomycin by either the oral or subcutaneous routes. In this experiment, it is seen that practically complete protection was afforded to all the treated mice. However, it should be noted that the mice that received 1 mg. of drug orally or 0.1 mg. subcutaneously were sick for 3 or 4 days, whereas the mice that received 5 mg. orally or 1 mg. subcutaneously never showed any signs of illness. Furthermore, it should be mentioned that although many of the control mice survived the infection, yet practically all control mice up through the 10^{-7} dilution were seriously ill for 16 to 19 days.

TABLE 5
LYMPHOGRANULOMA VENEREUM

INOCULATION OF YOLK-SAC SUSP		SURVIVAL RATIO OF MICE				CONTROLS
		TREATMENT*ONCE DAILY				
		ORAL DOSE		S C DOSE		
ROUTE	DILUTION**	5 MG	1 MG	1 MG	0.1 MG	
I C	10 ⁻³	5 / 5	5/5	5/5	4 / 4	1 / 5
	10 ⁻⁴	5/5	5/5	3/4	5/5	1 / 5
	10 ⁻⁵	5/5	5/5	4/4	4/4	3 / 5
	10 ⁻⁶	5/5	4/4	4/4	5/5	3 / 5
	10 ⁻⁷					2 / 5
	10 ⁻⁸					3 / 5
	10 ⁻⁹					5 / 5

* TREATMENT STARTED 24 HOURS AFTER INOCULATION AND CONTINUED FOR 5 DAYS

** 10^{-3} DILUTION = 645 LD₅₀

TABLE 6 shows the results obtained when mice were injected intravenously with approximately 4 LD₅₀ of murine typhus (*R. typhi*) and treated with graded doses of Aureomycin, starting treatment 24 hours post infection and continuing treatment for 5 days. It is seen that 0.1 mg. of Aureomycin is the least amount of drug affording complete protection although 0.01 mg. of the drug protected 6 out of 10 mice.

TABLE 7 shows the results obtained when mice were injected intravenously with approximately 4 LD₅₀ of murine typhus (*R. typhi*) and treated with Aureomycin by the subcutaneous route with 1 mg. per day for 5 days, starting treatment 24, 48, 72, and 96 hours post infection. It is seen that of the 20 active mice treated at 24 and 48 hours, all survived and none showed any signs of illness during the 18-day observation

TABLE 6

MURINE TYPHUS

INOCULATION OF YOLK-SAC SUSP		TREATMENT*	SURVIVAL RATIO OF MICE	
ROUTE	DILUTION	DAILY	TREATED	CONTROLS
		MG		
		1	10/10	
	1 5000**	0.5	10/10	
		0.1	10/10	
		0.01	6/10	
I V	1 5000**			0/10
	1 10 000			0/10
	1 20 000			5/10
	1 40 000			10/10 ..

* TREATMENT STARTED 24 HOURS AFTER INOCULATION
AND CONTINUED FOR 5 DAYS

** 1 5000 = 4 LD₅₀

TABLE 7

MURINE TYPHUS

INOCULATION OF YOLK-SAC SUSP		TREATMENT*	NO OF MICE IN GROUP		SURVIVAL
ROUTE	DILUTION	STARTED	AT START	TIME OF TREATMENT	RATIO
		24 HOURS	10	10 ACTIVE	10/10
		48 HOURS	10	10 ACTIVE	10/10
		72 HOURS	10	10 SICK	9/10
		96 HOURS	10	6 DEAD 4 SICK	4/4
		CONTROLS	13	9 DEAD 4 SICK	0/13

* 1 MG DAILY FOR 5 DAYS

** 1 5000 = 4 LD₅₀

period. With those treated at 96 hours, at which time 6 were dead and 4 were critically ill, all 4 treated mice survived. In contrast, 9 of 13 controls were dead at the end of 96 hours and all were dead by the 5th day. This experiment certainly demonstrates the marked beneficial therapeutic effect of Aureomycin in mice critically ill with murine typhus infection.

TABLE 8 shows the results obtained when mice were inoculated either intraperitoneally or intranasally with rickettsialpox (*R. akari*) and treated subcutaneously with Aureomycin for 12 days, starting 24 hours post infection. It is seen that all mice infected by the intraperitoneal route survived the infection without showing any signs of illness, regardless of whether they received 1 or 2 injections per day. In contrast, the 3 out of 10 controls that survived the infection were seriously ill for 7 to 10 days.

The mice infected intranasally and treated with Aureomycin likewise survived the infection without showing any signs of illness, whereas the 3 out of 9 controls that survived were seriously ill for 8 or more days.

TABLE 8
RICKETTSIALPOX

INOCULUM		TREATMENT*	SURVIVAL RATIO OF MICE	
ROUTE	DILUTION		TREATED	CONTROLS
I P	0.25 ML	DAILY	9/9	3/10
	YOLK-SAC	TWICE DAILY	9/9	
	SUSPENSION			
I N	0.05 ML AMNIOTIC SAC SUSP	DAILY	9/9	3/9

* TREATMENT STARTED 24 HOURS AFTER INOCULATION
AND CONTINUED FOR 12 DAYS

TABLE 9 shows the results obtained when mice were injected intraperitoneally with graded doses of yolk sacs infected with the Karp strain of scrub typhus (*R. tsutsugamushi*) and treated either orally or subcutaneously with Aureomycin, giving treatments on the 3rd, 4th, 7th, 8th, 11th, and 12th days post infection. It was found that 5 mg. of Aureomycin by the oral route completely protected all treated mice even though the infecting dose was as high as ten million lethal doses. One mg. of the drug by the oral route failed to protect any of the mice against ten million lethal doses, but protected 3 out of 5 mice against one million, and one-hundred thousand lethal doses, respectively, and completely protected 5 out of 5 mice against ten-thousand lethal doses. One mg. of Aureomycin given subcutaneously allowed 4 out of 5 mice to survive 10 million lethal doses, although all of these showed signs of mild illness for 4 days. The same amount of drug completely protected 5 out of 5 mice against one million lethal doses. In contrast, 39 of the 40 controls died from scrub typhus within 6 to 19 days post infection.

TABLE 9

SCRUB TYPHUS

INOCULATION OF YOLK-SAC SUSP		SURVIVAL RATIO OF MICE				
		TREATED*			CONTROLS	
		ORAL		S C	DAY DEATHS OCCURRED	
ROUTE	DILUTION	5MG	1MG	1MG		
I. P.	10^{-1}	5/5	0/5	4/5	0/5	6
	10^{-2}	4/4	3/5	5/5	0/5	8
	10^{-3}	5/5	3/5		0/5	11
	10^{-4}		5/5		0/5	13
	10^{-5}				0/5	15
	10^{-6}				0/5	14
	10^{-7}				1/5	14
	10^{-8}				0/5	19

* DRUG GIVEN ON 3, 4-7, 8-11 & 12TH DAYS AFTER INOCULATION

FIGURE 1 shows the results obtained when guinea pigs were injected intraperitoneally with graded doses of yolk-sac suspension infected with Rocky Mountain spotted fever (*R. rickettsii*), Bitterroot strain. Treatment with Aureomycin was started 48 hours post infection and consisted of 3 mg. subcutaneously once a day for 5 days. The relative temperature charts of the treated and control guinea pigs are shown in the slide. It should be pointed out that the charts represent the average temperatures of 5 guinea pigs in each of the treated groups and of 4 guinea pigs in each of the control groups. Ten out of 12 of the control guinea pigs showed scrotal swelling and 4 out of the 12 died. It is seen that complete protection against fever was afforded all treated guinea pigs against at least 2,500 infectious doses. Furthermore, none of the treated animals showed scrotal swelling or other signs of illness.

FIGURE 2 shows the results obtained when guinea pigs were injected intraperitoneally with 1 ml. of a 10 per cent brain suspension infected with epidemic typhus (*R. prowazeki*), Breinl strain. Treatment with Aureomycin was initiated on the 3rd, 4th, 5th, and 6th days post infection and consisted of 3 mg. subcutaneously once a day for 5 days. The average temperature charts of the guinea pigs in each group are shown in the chart. Groups 1, 2, and 5 each contained 4 guinea pigs, while Groups 3 and 4 had 5 guinea pigs each. Note that if treatment was initiated before fever appeared, the febrile reaction was completely suppressed. If treatment was initiated after fever appeared, the temperature was brought

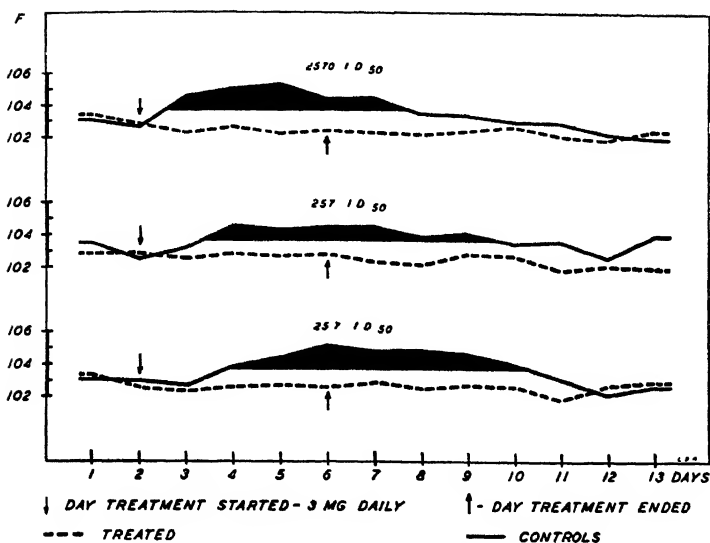


FIGURE 1. Treatment of guinea pigs inoculated intraperitoneally using yolk-sac suspension infected with Rocky Mountain spotted fever.

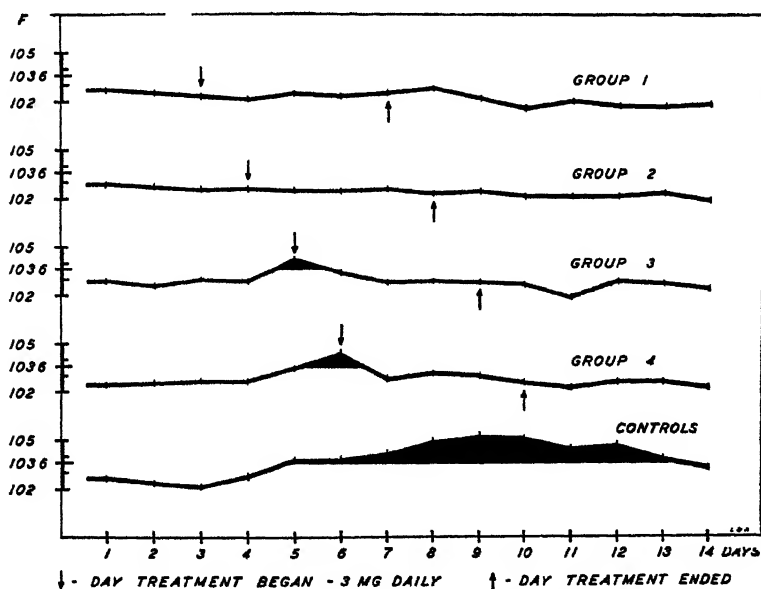


FIGURE 2. Treatment of guinea pigs inoculated intraperitoneally with 10% brain suspension infected with epidemic typhus.

back to normal within 24 to 48 hours. In contrast the control guinea pigs showed fever for an average of 7 days.

TABLE 10 shows the results obtained when guinea pigs were injected intraperitoneally with graded doses of spleen suspension infected with Q fever (*C. burnetii*, Nine Mile strain. Treatment with Aureomycin was started 24 hours post infection and consisted of 2.5 mg. subcutaneously twice a day for 8 days. It is seen that none of the 14 treated guinea pigs showed any febrile response or other signs of illness, whereas all 15 of the controls showed fever in dilutions as high as 1:1,000,000.

TABLE 10
Q FEVER

INOCULATION OF SPLEEN SUSPEN		TREATMENT*	GUINEA PIGS SHOWING TEMP	
ROUTE	DILUTION	TWICE DAILY	TREATED	CONTROLS
I P	10^{-2}	MG	0/3	3/3
	10^{-3}		0/3	3/3
	10^{-4}	2.5	0/2	3/3
	10^{-5}		0/3	3/3
	10^{-6}		0/3	3/3

*TREATMENT STARTED 24 HOURS AFTER INOCULATION
AND CONTINUED FOR 8 DAYS

FIGURE 3 shows the results obtained when guinea pigs were injected intraperitoneally with a 1:500 spleen suspension of Q fever, representing approximately 10,000 infectious doses. Treatment by the subcutaneous route with graded doses of Aureomycin was started 24 hours post infection and continued for 10 days.

The chart presents the average temperature of 4 guinea pigs in each group. The importance of the quantity of Aureomycin used in the daily dose is shown by the fact that Group 1 guinea pigs which received a daily dose of 3 mg. showed no febrile response; Group 2 which received 1 mg. daily showed low-grade fever for 48 hours; Group 3 which received 0.5 mg. daily showed the expected temperature rise but for a shortened period of time, namely 4 days; Group 4 which received 0.1 mg. daily showed no apparent difference from the control group and on the 4th day of fever the daily average dose of Aureomycin was increased to 4 mg. The temperature showed a drop to normal within 24 hours and stayed normal. The increased dosage was maintained for a total of 4 days. Controls receiving the same infectious inoculum showed continuous high-grade fever for 8 days, and controls receiving the same inoculum diluted to 1:10,000 times showed a lower temperature rise for 4 days

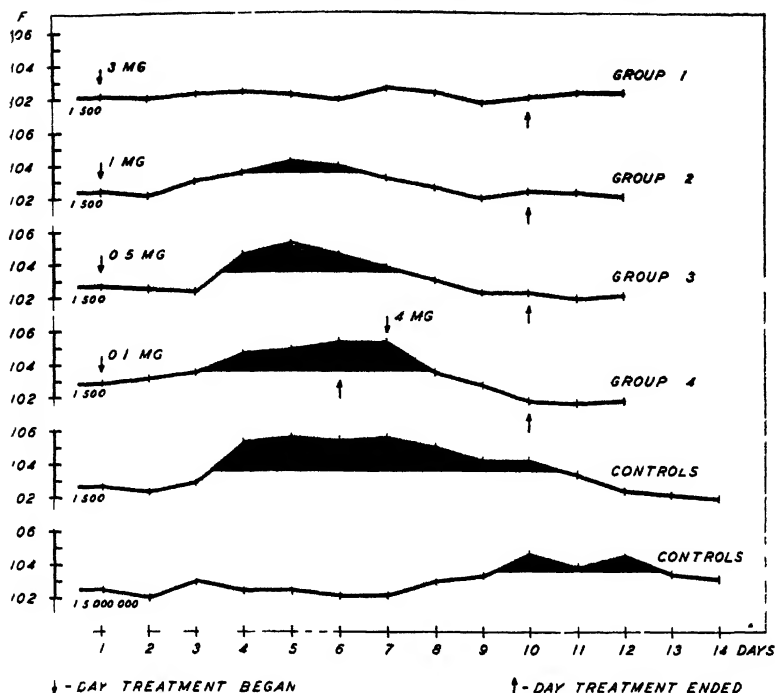


FIGURE 3. Treatment of guinea pigs inoculated intraperitoneally using spleen suspension infected with Q fever.

FIGURE 4 shows the results obtained when guinea pigs were injected intraperitoneally with a 1:500 spleen suspension of Q fever, representing approximately 10,000 infectious doses. Treatment with Aureomycin was started on the 3rd, 4th, 5th, 6th, and 7th days post infection and consisted of 3 mg. subcutaneously once a day for 3 days. The charts represent the average temperature of 4 guinea pigs in each group. It is apparent that all treated guinea pigs showed a return to normal temperature within 48 hours after the beginning of treatment, irrespective of the duration of fever prior to treatment. Essentially the same results were obtained in similar experiments carried out in guinea pigs infected with Rocky Mountain spotted fever and epidemic typhus.

In summary, from the data obtained thus far, the following general observations appear to be warranted:

1. The antibiotic Aureomycin has been shown to possess marked therapeutic activity against the viruses of the psittacosis-lymphogranuloma group and rickettsiae of the spotted fever, typhus fever, scrub typhus fever and Q fever groups in embryonated hens' eggs, mice, and guinea pigs.

2. Aureomycin shows no apparent *in vitro* activity.

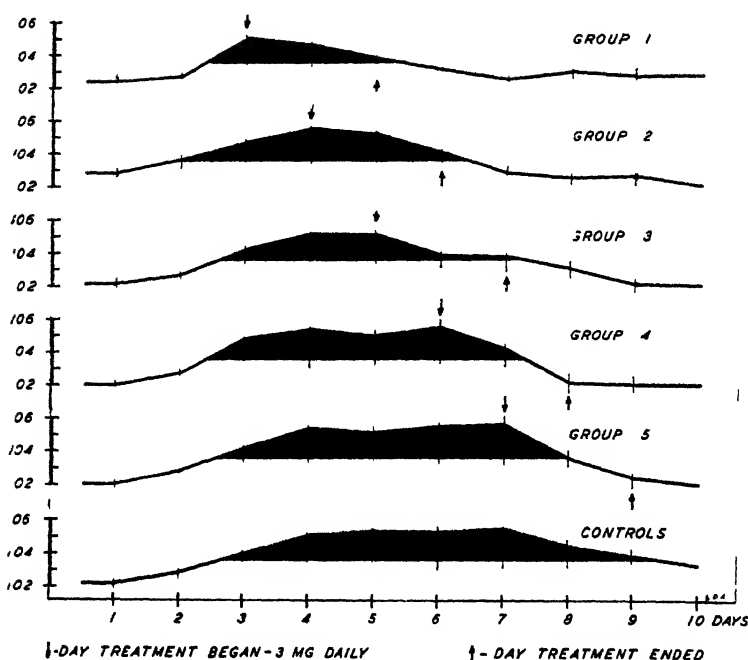


FIGURE 4. Treatment of guinea pigs inoculated intraperitoneally using spleen suspension infected with Q fever.

3. If guinea pigs receive a massive dose of infectious material and then are treated with Aureomycin before symptoms appear, no fever or other signs of illness develop, but antibodies in measurable degree are induced, and the animals are immune on subsequent rechallenge. On the other hand, if guinea pigs receive relatively small doses of infectious material and then are treated with Aureomycin before symptoms appear, no fever or other signs of illness develop, nor do antibodies always appear and the animals may or may not be immune on rechallenge. These findings indicate a quantitative relationship between the dosage of infecting agent and the dosage and time of administration of Aureomycin.

4. Regardless of the length of time fever has been apparent in guinea pigs infected with Rocky Mountain spotted fever, epidemic typhus or Q fever, the animals in most instances may be rendered afebrile within 48 to 72 hours by a single daily subcutaneous injection of 5 to 6 mg. of Aureomycin per kilogram body weight, for 3 to 5 days.

5. Aureomycin shows marked therapeutic activity orally, as well as parenterally, for mice infected intraperitoneally or intracerebrally with psittacosis or lymphogranuloma venereum viruses, or for mice infected intravenously or intranasally with rickettsialpox or murine typhus.

6. The question as to whether treatment with Aureomycin frees the tissues of the host from the infectious agent arises. This problem is under investigation, but the following comments may be made at the present time: Mice infected with large doses of psittacosis or lymphogranuloma venereum viruses and surviving following treatment with Aureomycin were found to harbor the infectious agent in their liver and brain, respectively, when sacrificed on the 17th day post infection. However, animals similarly infected and treated were found to be free of the infectious agent when tested on the 42nd day. In contrast, control animals infected with relatively small, nonfatal doses showed the infectious agent in their tissues at least through the 42nd day.

7. In other experiments carried out in our section, Aureomycin failed to show any therapeutic activity against the following viral infections: B strain of influenza, canine distemper, rabies, Newcastle disease, Venezuelan equine encephalomyelitis, and MEF-1 strain of poliomyelitis. Preliminary experiments with mumps virus infected chick embryos showed that Aureomycin reduced or completely inhibited the production of virus as measured by the hemagglutinating activity of the virus (allantoic fluid) for chicken red cells, but had no effect on the hemagglutinating activity of the virus *in vitro*, and apparently had no appreciable influence on the rate of multiplication and infective titer of the mumps virus as measured by infectivity tests carried out in embryonated hens' eggs.

Bibliography

- ANDREWES, C. H., M. VAN DEN ENDE, H. KING, & J. WALKER. 1944. Substances chemotherapeutically active against typhus rickettsiae. *Lancet* 246: 777.
- ANDREWES, C. H., H. KING, & M. VAN DEN ENDE. 1943. Chemotherapeutic experiments with viruses of influenza A, lymphogranuloma venereum and vaccinia. *J. Path. & Bact.* 55: 173-181.
- ANDREWES, C. H., H. KING, & J. WALKER. 1946. Chemotherapeutic action of dyes in typhus infection of mice. *British J. Pharm. & Chemotherapy* 1: 15-20.
- ANIGSTEIN, L., & M. N. BADER. 1945. Para-aminobenzoic acid—its effectiveness in spotted fever in guinea pigs. *Science* 101: 591-592.
- CUTTING, W. C., R. H. DREIBACH, R. M. HALPERN, E. A. IRWIN, D. W. JENKINS, F. PROESCHER, & H. B. TRIPI. 1947. Chemotherapy of virus infections. *J. Immunol.* 57: 379-380.
- EATON, M. D., A. VAN ALLEN, & A. WIENER. 1947. Action of acridines on agents of psittacosis-lymphogranuloma group. *Proc. Soc. Exp. Biol. & Med.* 66: 141-146.
- EATON, M. D., T. F. DOZOIS, A. VAN ALLEN, V. L. PARISH, & S. SCHWALM. 1948. Studies on inhibition of the agents of murine and feline pneumonitis by penicillin. *J. Immunol.* 58: 251-267.
- EATON, M. D., & V. L. HANFORD. Effect of the host in action of sulfonamides on elementary body agents of murine and feline pneumonitis. *Proc. Soc. Exp. Biol. & Med.* 59: 63-66.
- EHRLICH, J., Q. R. BARTZ, R. M. SMITH, & D. A. JOSLYN. 1947. Chloromycetin, a new antibiotic from a soil actinomycete. *Science* 106: 417.
- FLORMAN, A. L., A. B. WEISS, & F. E. COUNCIL. 1946. Effect of large doses of streptomycin and influenza viruses on chick embryos. *Proc. Soc. Exp. Biol. & Med.* 61: 16-18.
- FOX, J. P., & O. L. PETERSON. 1948. The anti-rickettsial effect of thionine dyes. II. On the mode of action of the thionine dyes in combatting experimental in-

- fections of mice with *Rickettsia oryntalis* and *Rickettsia mooseri*. *J. Immunol.* **58**: 299-321.
- GREEN, R. H., A. F. RASMUSSEN, & J. E. SMADEL. 1945. Chemoprophylaxis of experimental influenza infections in eggs. *Pub. Health Rep.* **61**: 1401-1403.
- HAMILTON, H. L. 1945. Effect of P-aminobenzoic acid on growth of different rickettsiae and elementary bodies. Its mode of action. *Proc. Soc. Exp. Biol. & Med.* **59**: 220-226.
- HAMILTON, H. L., H. PLOTZ, & J. E. SMADEL. 1945. Effect of para-aminobenzoic acid on the growth of typhus rickettsiae in the yolk sac of the infected chick embryo. *Proc. Soc. Exp. Biol. & Med.* **58**: 255-262.
- HAMRE, D., & G. RAKE. 1947. Studies on lymphogranuloma venereum. V. The action of some antibiotic substances and sulfonamides *in vitro* and *in vivo* upon the agents of feline pneumonitis and lymphogranuloma venereum. *J. Infect. Dis.* **81**: 175-190.
- HUEBNER, R. J., G. A. HOTILE, & E. B. ROBINSON. 1948. Action of streptomycin in experimental infection with Q fever. *Pub. Health Rep.* **63**: 357-362.
- JONES, H. P., G. RAKE, & C. M. MCKEE. 1941. Chemotherapy of lymphogranuloma venereum with sulfonamide drugs. *Proc. Soc. Exp. Biol. & Med.* **48**: 318-323.
- JONES, H., G. RAKE, & B. STEARNS. 1945. Studies on lymphogranuloma venereum. III. The action of the sulfonamides on the agent of lymphogranuloma venereum. *J. Infect. Dis.* **76**: 55-69.
- KRAMER, S. D., H. A. GEER, & D. A. SZOBEL. 1944. Chemoprophylactic and therapeutic action of wide variety of chemical compounds on two neurotropic virus infections in mice. *J. Immunol.* **49**: 273-314.
- LEY, H. L., JR., J. E. SMADEL, & T. T. CROCKER. 1948. Administration of Chloromycetin to normal human subjects. *Proc. Soc. Exp. Biol. & Med.* **68**: 9-12.
- MCKINSTRY, D. W., & E. H. READING. 1944. Studies on the chemotherapy of experimental virus infections. I. The effect of certain pyrimidine derivatives on experimental murine poliomyelitis. *J. Franklin Institute* **237**: 422-431.
- McLIMANS, W. F., & C. C. GRANT. 1947. Therapy of experimental tsutsugamushi disease (scrub typhus). *Science* **105**: 181-182.
- MEIKLEJOHN, G., J. C. WAGNER, & G. W. BEVERIDGE. 1946. Studies on the chemotherapy of viruses in the psittacosis lymphogranuloma group. I. Effect of penicillin and sulfadiazine on ten strains in chick embryo. *J. Immunol.* **54**: 1-8.
- MORGAN, H. R., D. A. STEVENS, & J. C. SNYDER. 1947. Effect of streptomycin on growth of rickettsiae in eggs. *Proc. Soc. Exp. Biol. & Med.* **64**: 342-345.
- PARKER, R. F., & H. W. DIFENDORF. 1944. Effect of penicillin on certain viruses. *Proc. Soc. Exp. Biol. & Med.* **57**: 351-354.
- PAYNE, E. H., J. A. KNAUDT, & S. PALACIOS. 1948. Treatment of epidemic typhus with chloromycetin. *J. Trop. Med. & Hyg.* **51**: 68-71.
- PETERSON, O. L., & J. P. FOX. 1947. The anti-rickettsial effect of thionine dyes. I. The use of methylene blue and toluidine blue to combat experimental tsutsugamushi disease (scrub typhus). *J. Exp. Med.* **85**: 543-558.
- RAKE, G., & D. M. HAMRE. 1944. Action of sulfonamide on toxins of agents of lymphogranuloma psittacosis group. *Proc. Soc. Exp. Biol. & Med.* **55**: 90-91.
- RAKE, G., & H. JONES. 1946. The toxicologic and therapeutic action of penicillin on mice infected with the agent of lymphogranuloma venereum. *Am. J. Syph., Gon. & Ven. Dis.* **30**: 242-246.
- RAVENEL, S. F. 1947. Para-aminobenzoic acid therapy of Rocky Mountain spotted fever. Outline of a comprehensive plan of treatment with a report of five cases. *J. A. M. A.* **133**: 989-994.
- SMADEL, J. E., & E. G. JACKSON. 1947. Chloromycetin, an antibiotic with chemotherapeutic activity in experimental rickettsial and viral infections. *Science* **106**: 418-419.
- SMADEL, J. E., & E. G. JACKSON. 1948. Effect of Chloromycetin on experimental infection with psittacosis and lymphogranuloma venereum viruses. *Proc. Soc. Exp. Biol. & Med.* **67**: 478-483.
- SMADEL, J. E., A. P. LEON, H. L. LEY, JR., & G. VARELA. 1948. Chloromycetin in the treatment of patients with typhus fever. *Proc. Soc. Exp. Biol. & Med.* **68**: 12-19.
- SMADEL, J. E., J. C. SNYDER, E. B. JACKSON, J. P. FOX, & H. L. HAMILTON. 1947.

- Chemotherapeutic effect of acridine compounds in experimental rickettsial infections in embryonated eggs. *J. Immunol.* **57**: 155-171.
- SMITH, R. M., D. A. JOSLYN, O. M. GRUHZIT, I. W. MCLEAN, JR., M. A. PENNER, & J. EHRLICH. 1948. Chloromycetin. Biological studies. *J. Bact.* **55**: 425-448.
- SNYDER, J. C., A. YEOMANS, D. H. CLEMENT, E. S. MURRAY, C. J. D. ZARAFONETIS, & N. A. TIERNEY. 1947. Further observations on the treatment of typhus fever with para-amino-benzoic acid. *Ann. Intern. Med.* **27**: 1-27.
- STEINHAUS, E. A., & R. R. PARKER. 1943. Experimental Rocky Mountain spotted fever: Results of treatment with certain drugs. *Pub. Health Rep.* **58**: 351-353.
- STUART-HARRIS, C. H., M. VAN DEN ENDE, G. WHITE, W. L. HAWLEY, J. S. F. NIVEN, with A. M. BEGG, F. FULTON, F. HAMILTON, & C. C. THOMAS. 1946. Chemotherapeutic and other studies of typhus. *Med. Res. Council. Sp. Rep.*: 255.
- TIERNEY, N. A. 1946. Effect of para-aminobenzoic acid in Tsutsugamushi disease. *J. A. M. A.* **131**: 280-285.
- WISEMAN, R. W., G. MEIKLEJOHN, D. B. LACKMAN, J. C. WAGNER, & G. W. BEVERIDGE. 1946. Studies on the chemotherapy of viruses in the psittacosis lymphogranuloma group. II. Effect of penicillin and sulfadiazine on seven strains in mice. *J. Immunol.* **54**: 9-16.
- ZARAFONETIS, C. J. D., J. C. SNYDER, & E. S. MURRAY. 1946. Immunity following para-aminobenzoic acid therapy in experimental Tsutsugamushi disease (scrub typhus). *Proc. Soc. Exp. Biol. & Med.* **61**: 240-242.

AUREOMYCIN—A NEW ANTIBIOTIC WITH ANTIRICKETTSIAL PROPERTIES: ITS EFFECT ON EXPERIMENTAL SPOTTED FEVER AND EPIDEMIC TYPHUS*

By LUDWIK ANIGSTEIN, DOROTHY M. WHITNEY, AND JOE BENINSON
*Rickettsial Research Laboratory, Department of Preventive Medicine and Public Health,
University of Texas Medical Branch, Galveston, Texas*

THE new antibiotic recently isolated by the Lederle Laboratories Division, American Cyanamid Company, from the mold *Streptomyces aureofaciens* Duggar,¹ and named Aureomycin, has an unusual spectrum of activity covering a wide range of pathogenic organisms including rickettsiae and certain viruses.^{2, 3}

The present report concerns the activity of aureomycin in two rickettsial diseases, namely, spotted fever and epidemic typhus. The observations are based on the clinical, pathological and immunological responses of the infected guinea pig under treatment with the antibiotic.

Materials and Methods

The highly virulent strain of Rocky Mountain spotted fever obtained from the Rocky Mountain Laboratory induces in guinea pigs clear-cut symptoms, characteristic pathological lesions, and a fatality rate of 100 per cent. The well known Breinl strain of epidemic typhus was supplied by Dr. J. C. Snyder of Harvard in cotton rats from which it has been established in guinea pigs. Whereas the incubation period in experimental spotted fever seldom exceeds 4 days, the typhus infected guinea pig reacts with abrupt fever after 7-8 days. In both infections, the fever runs between 40° and 41° C., usually of a continuous type over a period of 5-7 days. There is no fatality among guinea pigs due to typhus infection.

Transfers of spotted fever were made by the intraperitoneal or subcutaneous routes, injecting 1 ml. of a 10 per cent splenic emulsion derived from a frozen spleen of an infected guinea pig and suspended in skim milk. A fairly homogeneous suspension was secured by filtration through double surgical gauze. For typhus, similarly treated brain tissues served as the source of infection. Constant volumes of 1 ml. of the infectious material were used in all experiments.

Aureomycin hydrochloride was supplied by the Lederle Laboratories in two forms: vials containing 20 mg. of the golden-yellow powder, readily soluble in water for parenteral use; capsules each containing 50 mg. of the antibiotic with a small amount of impurities present. The latter form

* This study is supported by a grant from the Lederle Laboratories Division, American Cyanamid Company

is designated for oral use only. Both modes of administration were used in the present study.

Guinea pigs of both sexes, 400-500 gm. body-weight were used and maintained prior and during the experimental period in an air-conditioned laboratory (76-78° F). Daily rectal temperatures of the animals were recorded for at least 18 days during the infection.

Experiments and Results

Following the unpublished reports of Lederle Laboratories, the general trend of the present study was to establish an efficient mode of administration of aureomycin which will protect the guinea pig from disease. This includes the minimum effective dosage, the frequency of administration, and the duration of the treatment. Emphasis was placed on the initial phase of the infection covering the entire or partial incubation period. Except for two series of experiments when the initial dose of the drug followed shortly after the infection, in all other series a delay of 24 or 48 hours in the treatment was adopted. Furthermore, attempts were made to establish the longest delay in drug administration, to counteract the disease and to inhibit febrile reactions while the disease was in progress. Consequently, clinical prophylaxis combined with curative measures were the objectives of the present study.

The activity of aureomycin in experimental rickettsial infections has been first investigated by the Lederle Laboratories. Their unpublished progress reports included spotted fever and Q fever in guinea pigs; rickettsialpox and murine typhus in mice. The tests with spotted fever showed that a daily subcutaneous injection of 4 mg. of aureomycin, 24 hours after infection and over a period of 7 days, protected guinea pigs against the infection. The total dosage in these experiments amounted to 28 mg. or 60 mg. kg. per guinea pig. Similar results were recorded in guinea pigs infected with Q fever and treated with 5 mg. daily for a total of 8 days (90 mg. kg.). The tests against murine typhus in mice injected daily for 5 days with 0.1 mg. aureomycin resulted in protection against otherwise lethal typhus infection. No untoward or toxic effects of this dosage were reported by the Lederle Laboratories.

These encouraging results, offering a wide field for further developments toward an optimal chemotherapy, stimulated us to elaborate and analyze the antirickettsial activity of this antibiotic.

Spotted Fever. Initial tests were performed on a series of 8 guinea pigs subcutaneously infected with spotted fever and treated with aureomycin by subcutaneous injections of 2 mg. daily for 7 days beginning with the day of infection (FIGURE 1). As compared with the controls showing a typical course of spotted fever followed by death, the response of all test animals was significant: 4 of them remained afebrile during 18 days' observation, 2 others reacted with moderate or abortive fevers on the 13th day, and one responded on the same day with slight fever of one-day

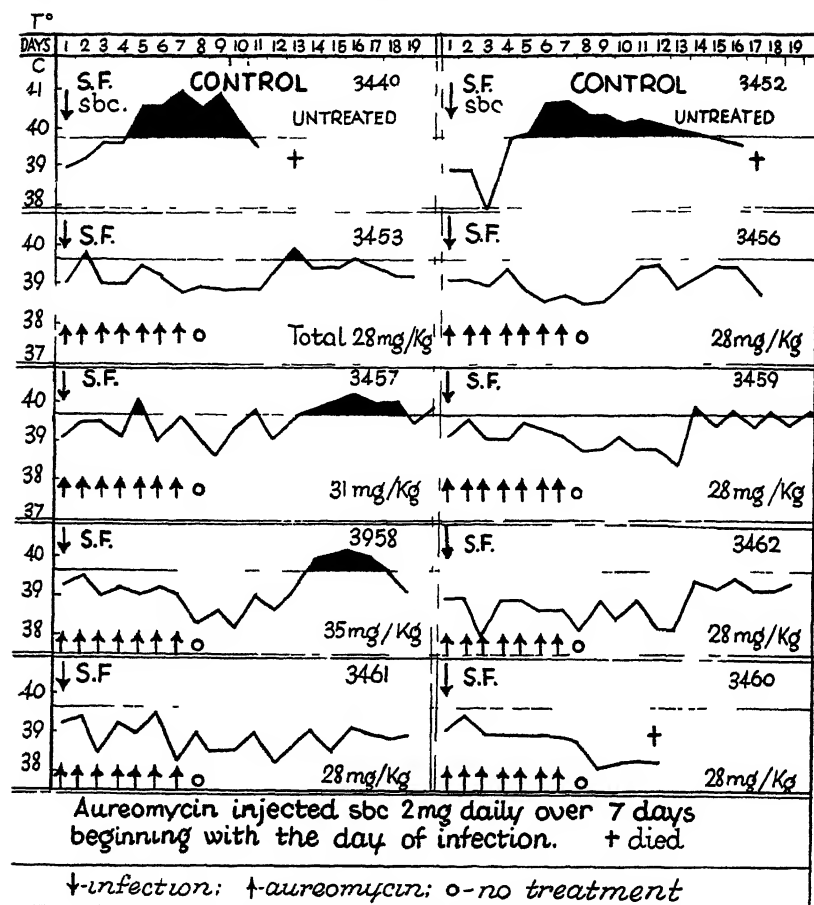
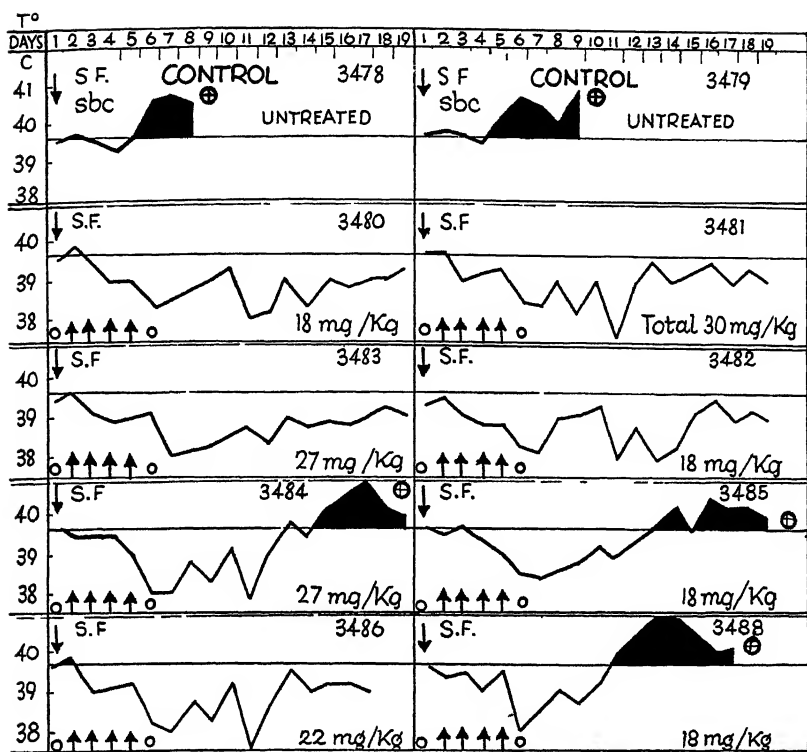


FIGURE 1. Spotted fever in guinea pigs treated with aureomycin

duration. With the exception of one animal which died probably of diarrhea and showed hemorrhages at the site of drug injection, all other guinea pigs recovered (FIGURE 1). It was concluded from this initial series that aureomycin had a definite suppressive effect on the course of the infection in preventing the clinical symptoms or attenuating the disease after a protracted incubation.

All recovered guinea pigs of this series were reinoculated by subcutaneous route with the routine heavy dose of spotted fever, one month after the initial infection. None of the animals showed immunity.

In the next series, the administration of the drug (2 mg. daily) was initiated 24 hours after infection and continued for 4 days during incubation (FIGURE 2). Thus, the total amount of the antibiotic was reduced



Aureomycin injected sbc 2mg. daily over 4 days beginning with 2nd day of infection. ⊕-killed

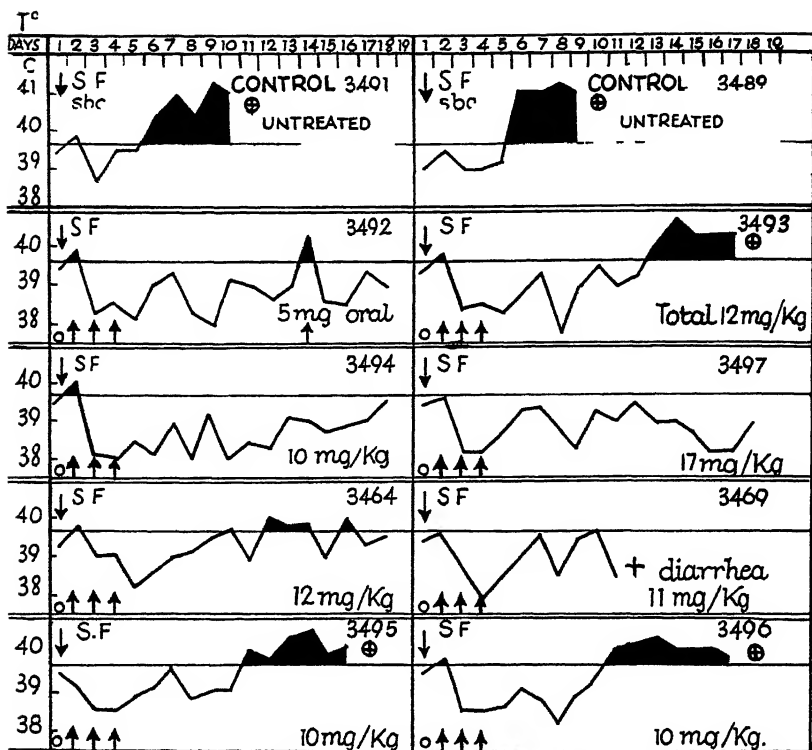
↓-Infection; ↑-Aureomycin; ○-No treatment

FIGURE 2. Spotted fever in guinea pigs treated with aureomycin.

from the average of 30 mg./kg. to 22 mg./kg. Of 8 test animals, 5 remained entirely afebrile whereas 3 developed spotted fever after 10-14 days, contrary to the controls which came down with fever after the usual 4 days incubation. Consequently, despite the 24 hour delay in drug administration and the considerable shortening of the treatment, the clinical response to the infection was prevented in 62 per cent of animals in this series (FIGURE 2).

No immunity to reinoculation with spotted fever was shown by the afebrile guinea pigs.

A further reduction in dosage was made in the third series of 8 guinea pigs infected with spotted fever 24 hours prior to drug administration, and injected with 2 mg. daily for 2 days and with 1 mg. on the fourth day.



Aureomycin injected sbc 2mg daily on the 2nd & 3rd day and 1mg on the 4th day of infection.

↓ Infection, ↑ Aureomycin ⊕ Killed, o-No treatment

FIGURE 3 Spotted fever in guinea pigs treated with aureomycin

A total amount of 5 mg. per guinea pig, or an average of 12 mg. kg. were used during the experiment. As seen from charts of FIGURE 3, moderately severe fever reactions (as compared with the controls) were recorded in 3 guinea pigs after a 10–12 day incubation; two other animals remained afebrile and 2 showed abortive and slight fevers. One of them (3492) deserves attention because it was treated with 5 mg. of aureomycin orally on the 14th day after an abrupt fever attack (40.4° C.) was recorded. On the following morning, the temperature subsided as abruptly as it rose the day before.

The prompt response to a single dose of aureomycin given orally justified further exploration of this procedure. For this purpose, a small group of 4 guinea pigs (plus 2 controls) were each given 10 mg. aureomycin in aqueous solution (from capsules) orally 1 hour before the intra-

peritoneal infection with spotted fever. The same single dose was repeated the following day after which the treatment was discontinued. Both untreated controls came down with high fever after 4 days, whereas 2 test animals showed simultaneous slight temperature elevations on the 9th day. They were then given a single oral dose of 5 mg. aureomycin in distilled water after which the temperature was normal on the following morning. Guinea pig 3508 showing a second rise of fever was injected with 3 mg. of the drug. Another drop of temperature was recorded. The third member of the group (3511) remained afebrile until the 13th day when an abrupt rise to 40.2° C. was noted. In this case, high fever persisted despite 8 mg. of the antibiotic injected during the first 2 days of fever.

The initial treatment by the oral route was continued in a series of 10 guinea pigs with the dosage of 5 mg./kg. beginning with the second day after infection and continued for the following 2 days (FIGURE 4). As compared with the controls, which all reacted with abrupt and high fevers on the fourth day, the course of the disease in all test animals was greatly modified. First, in most of the cases, a sheer drop in body temperature to subnormal levels accompanied the intake of the antibiotic. This was followed by an almost simultaneous but moderate rise of fever on the 6th or 7th day after infection. Except for guinea pig 3533 which developed a protracted and atypical course of fever, in all 9 animals the first attack responded promptly to a single injection of 1-2 mg. aureomycin. In some animals, the temperature remained normal, despite the withdrawal of the antibiotic for 6 days, until the afebrile period was interrupted by a secondary and longer fever attack on the 10th or 14th day (FIGURE 4). Treatment was resumed if and when indicated by fevers. As compared with the prompt response of the initial symptoms, the treatment of the secondary fever attacks was as a whole less effective. Except for 2 guinea pigs which were killed, all other test animals recovered.

As to the character of the fever charts, a relapsing type can be recognized beginning with an abortive attack after a prolonged incubation period. An afebrile period of several days duration followed and was terminated by fever of varying intensity. Whether the quiescent period is due to the activities of the antibiotic *per se* or is correlated with the interaction of immune factors cannot be answered without speculation. In this regard, the significant level of aureomycin for at least 24 hours after a single injection, as demonstrated in blood of patients with lymphogranuloma venereum, should be emphasized.⁸ The effect of aureomycin in our series becomes more significant considering the severe challenge of the intraperitoneal route of the infection resulting in the fulminating disease and death in the controls.

The next inquiry was to establish the longest possible delay in the protection against spotted fever by aureomycin. Since, in the great majority of the control guinea pigs, the length of the incubation period was 72 hours, it was decided to initiate treatment 48 hours after the

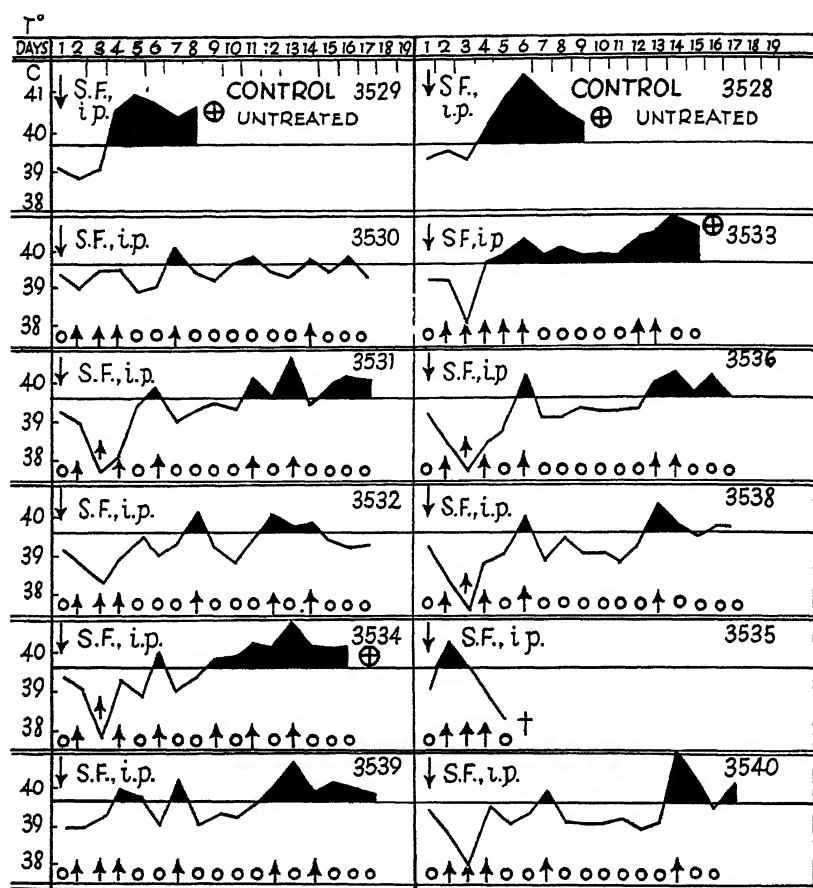


FIGURE 4. Spotted fever in guinea pigs treated with aureomycin.

infection. A series of 12 guinea pigs was infected intraperitoneally with the usual dose of splenic suspension. The untreated control guinea pigs reacted with abrupt fevers after 72 hours and ran a typical severe course of spotted fever.

To all 12 test animals, a dose of 5 mg. aureomycin in distilled water was given orally, 48 hours after infection. An "intermittent" treatment

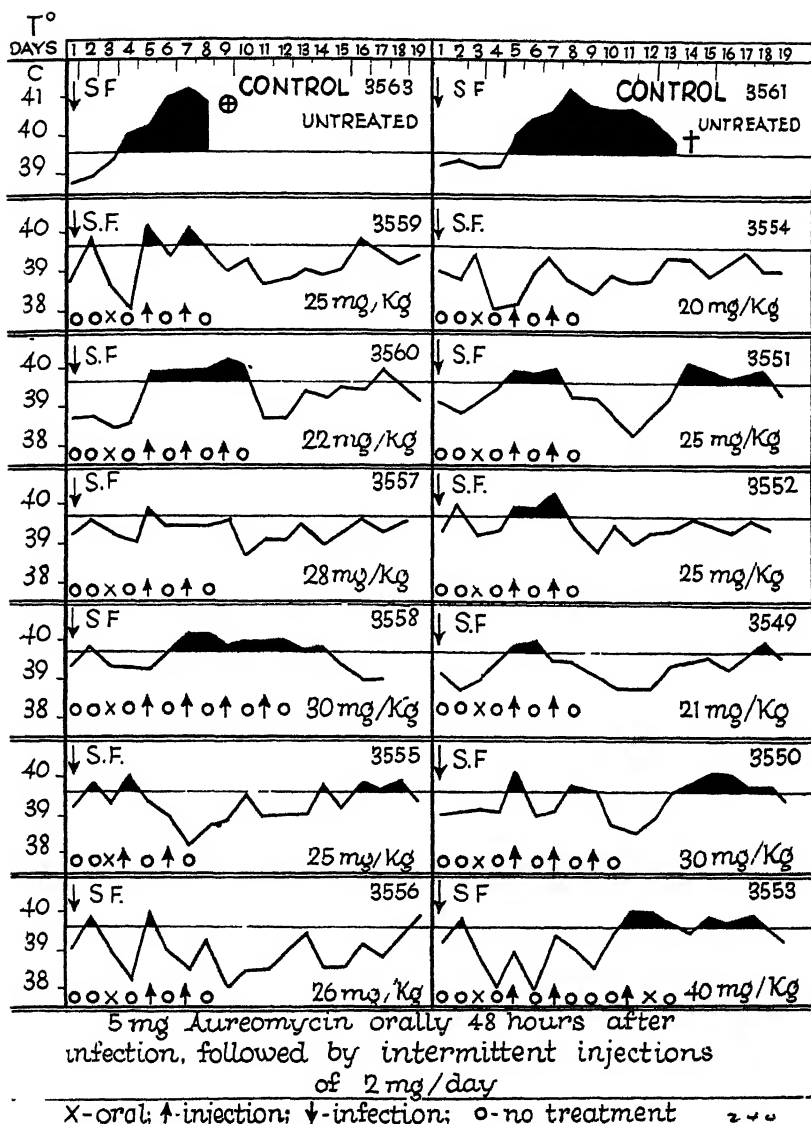


FIGURE 5. Spotted fever in guinea pigs treated with aureomycin

was then adopted by injecting 2 mg. of aureomycin every third day (FIGURE 5).

Except for 4 animals which showed more protracted fevers and received a total of 3 or 4 intermittent injections, the remaining 8 guinea pigs of this series were injected with 2 mg. aureomycin each on the 5th

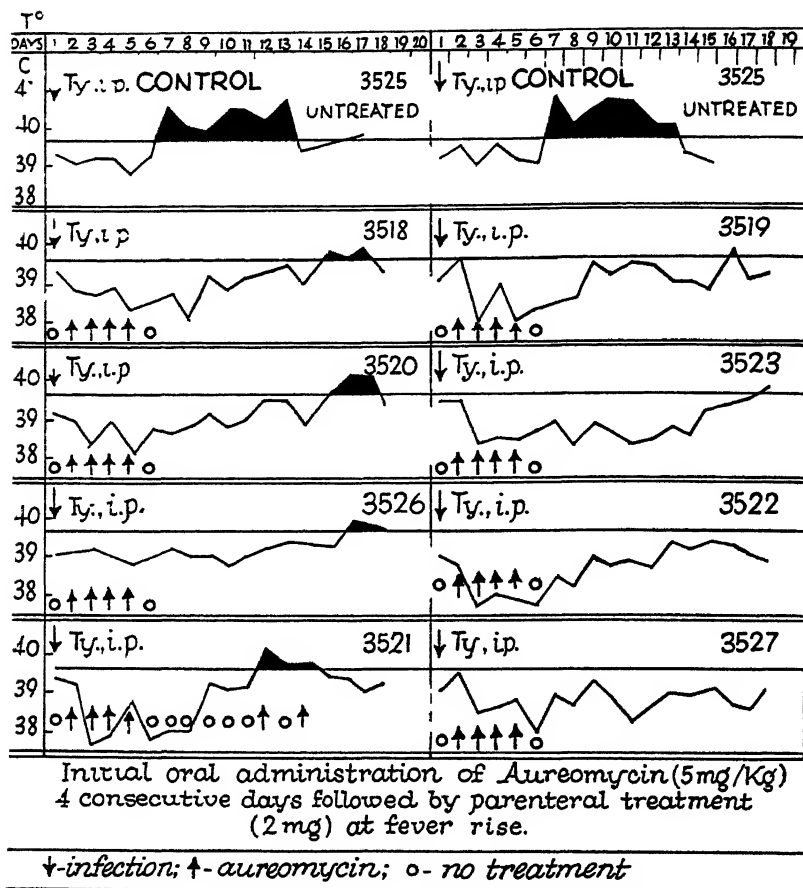
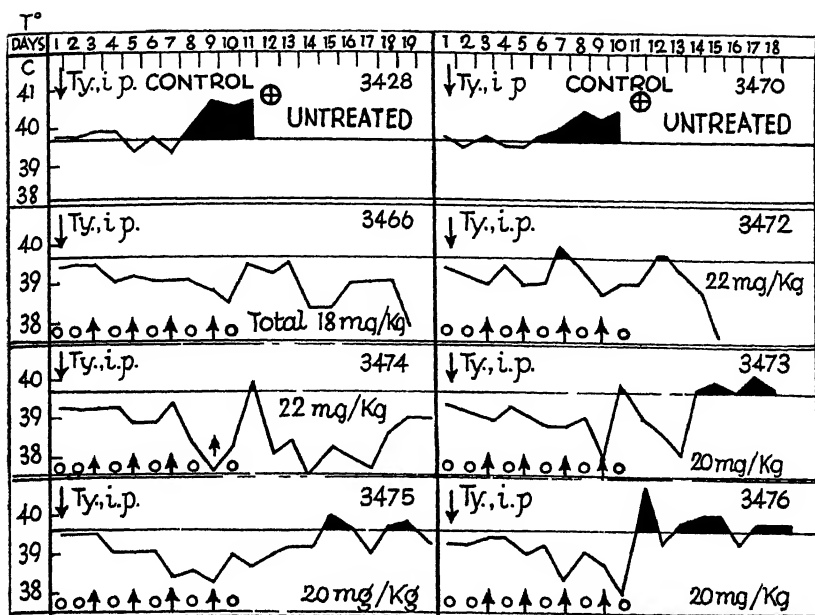


FIGURE 6. Epidemic typhus in guinea pigs treated with aureomycin

and 7th day of infection (FIGURE 5). Apart from guinea pig 3554, which remained completely afebrile after a total of 10 mg. aureomycin, a wide variety of fever reactions was observed in the remaining test animals.

In some cases, slight temperature elevations of 1, 2 or 3 days duration at the early stages of the infection were noted, while protracted and moderate fevers were recorded in guinea pigs Nos. 3558 and 3560. Definite relapses separated from initial attacks by afebrile periods were also observed in this series. On the whole, and as compared with the controls, a significant suppression of clinical symptoms was achieved when the initial dose was given 48 hours after the infection followed by a short intermittent parenteral treatment. All treated animals recovered, gradually regaining the heavy loss in body weight (up to 25 per cent), particularly by the end of the first week.



Intermittent treatment with 2 mg Aureomycin
 injected every 48 hrs. from 3rd to 9th day of infection (i.p.)

↓ -infection; ↑ -aureomycin; ○ -no treatment, ⊕ -killed

FIGURE 7. Epidemic typhus in guinea pigs treated with aureomycin.

Epidemic Typhus. The characteristic and consistent fever reactions of guinea pigs infected with the well-established strain of epidemic typhus were the criteria for evaluating the effect of aureomycin on this rickettsial infection.

In a series of 8 guinea pigs, oral administration of aureomycin was initiated 24 hours after intraperitoneal infection with typhus brain suspension (FIGURE 6). The dosage of 5 mg./kg. guinea pig per day for 4 consecutive days was given, after which the treatment was discontinued. In the 4 untreated controls, a sharp rise of fever was recorded on the 7th day, and remained on a high level for the next 7 days. In sharp contrast, 4 test animals remained afebrile during the 18-day observation, while 4 showed slight or moderate fevers on the 15th or 16th day. As a whole, typhus fever was either entirely suppressed or reduced to abortive negligible attacks after a considerably prolonged incubation.

The mode of treatment of the next series of 6 guinea pigs consisted in parenteral administration only, given at intervals every third day and beginning 48 hours after i.p. infection (FIGURE 7). Daily subcutaneous

injections were given on the 3rd, 5th, 7th, and 9th day of the infection, whereupon the treatment was discontinued.

Irregular intermittent fever attacks developed in 3 guinea pigs after a protracted incubation while the other 3 guinea pigs remained afebrile.

No reports by other authors on the action of aureomycin on epidemic typhus were available. However, murine typhus in mice was investigated by the Lederle Laboratories. The results of Wong and Cox indicate that aureomycin affords complete protection to a lethal dose of murine typhus in mice.

Summary and Discussion

It is evident from the results of the present study that aureomycin is provided with potent antirickettsial properties as previously reported by the Lederle Laboratories. In their early progress report, a daily injection of 4 mg. aureomycin per guinea pig given over a period of 7 days (total 62 mg. kg.) protected the infected animals from spotted fever. In our experience, a protection against spotted fever was achieved with less than half or even with one-third of this dosage when treatment was initiated 24 or 48 hours after i.p. infection. It was also found that daily treatment for 7 consecutive days had no appreciable advantage over intermittent administration every third day with lesser amount of the antibiotic. The various modifications of the treatment still in progress seem to indicate that the initial phase of chemotherapy has a decisive influence, not only on the prolongation of the incubation period, but also on the ultimate complete or partial suppression of clinical symptoms. In the latter case, they are of the abortive type and can be checked in most cases by single injections of 1-2 mg. aureomycin. Late relapsing attacks of spotted fever are less responsive to treatment and point to a latent character of the suppressed infection. With the exception of a few guinea pigs which developed diarrhea and died, all test animals recovered, contrary to the 100-per-cent fatality rate of the untreated controls.

In typhus infected guinea pigs, still more consistent results were achieved. With the exception of a few abortive and late fever reactions, the protection in the great majority of cases was complete. This was obtained by initial oral treatment 24 hours after injection with 5 mg./kg. daily per animal for 4 consecutive days. Similar beneficial effects of aureomycin resulted from parenteral treatment beginning 48 hours after infection and then at 48-hour intervals for 4 days.

In studying the temperature curves of infected guinea pigs, attention was paid not only to the febrile reactions, but also to the afebrile periods. In the light of our previous experience with PABA, it was assumed that the effect of aureomycin is due, like PABA, to its rickettsiostatic activity resulting in the inhibition of rickettsiae and in the suppression of symptoms. A disturbance of the balance between the inhibited rickettsiae and the defensive mechanism of the infected host would then explain the

flares of abortive attacks and relapses. Consequently, an immunity to the homologous rickettsiae should be afforded by the subclinical infection as was the case in PABA treated guinea pigs. However, in preliminary experiments, aureomycin-treated guinea pigs showed no immunity to reinoculation with massive infective doses of spotted fever. Furthermore, spleen of afebrile guinea pigs under antibiotic treatment when injected into normal ones induced no reactions. These results are comparable with the early findings of the Lederle Laboratories on murine typhus in mice in which, after treatment, no latent infection was found.

This intricate phase of the immune phenomena in chemotherapy deserves further investigation, which is in progress.

Addendum

While this paper was submitted for publication, the series illustrated in FIGURE 5 was reinoculated with spotted fever 4 weeks after the initial infection. Splenic suspension in 1:100 dilution was used for these tests instead of 1:10 as previously injected. Complete immunity was established in every guinea pig tested, regardless of its previous clinical responses. The immunity developed by afebrile guinea pigs is significant as it indicates a symptomless (inapparent) infection due to the suppressive, rickettsiostatic action of aureomycin.

References

1. DUGGAR, B. M. 1948. *Ann. N. Y. Acad. Sci.* 51(2): 177-181.
2. WONG, SAM C., & HERALD R. COX. 1948. *Ann. N. Y. Acad. Sci.* 51(2): 290-305.
3. WRIGHT, L. T., M. SANDERS, *et. al.* 1948. *Ann. N. Y. Acad. Sci.* 51(2): 318-330.

THE TREATMENT OF LYMPHOGRANULOMA VENEREUM AND GRANULOMA INGUINALE IN HUMANS WITH AUREOMYCIN*

By L. T. WRIGHT, M. SANDERS, M. A. LOGAN, A. PRIGOT,
AND L. M. HILL

*Surgical Department of Harlem Hospital, Department of Hospitals, New York City,
and the Ophthalmological Department, College of Physicians and Surgeons,
Columbia University, New York, N. Y.*

IN a preliminary study which is in press,¹ we have reported results in twenty-five cases of lymphogranuloma venereum treated with Aureomycin. This study was conducted between January 22nd, 1948, and April 24th, 1948. The results obtained with this new therapeutic agent were so superior to all other forms of therapy² that had been used by us that we did not hesitate to continue its use. In the earlier paper, it was stated that Aureomycin was used for the first time in human beings. Apparently we were dealing with an antibiotic that could cure this virus infection in humans.

The present paper confirms our previous observations. It is a report of ten additional cases of lymphogranuloma venereum, treated with Aureomycin since our original article was sent in for publication. It also includes a follow-up report on fourteen of our original twenty-five cases, and a preliminary report on the use of Aureomycin in three cases of granuloma inguinale.

Lymphogranuloma Venereum

To date, the total number of patients to whom we have administered Aureomycin, as the sole form of specific therapy, is thirty-five, and it should be stated that they were all hospitalized. This was done so that we could be sure that the patients received an adequate amount of the drug under controlled conditions and, also, so that we could properly evaluate the results.

Most patients received 10 to 40 mg. of Aureomycin daily, by intramuscular injection. The drug was dissolved in 2 cc. of normal saline solution. In our previous series, a mild, hypochromic anemia developed when we used a special diluent. In all cases, the anemia was corrected by the administration of folic acid and iron (Folvron). In some of our earlier cases, we used folic acid and iron as an anemia preventive, while in this group we found this precaution unnecessary.

Because of a pre-existing anemia, two patients in this series received folic acid and iron while being treated with Aureomycin. We feel that the

* We are grateful to the late Dr. Y. Subbarow and his associates at the Lederle Laboratories Division, American Cyanamid Company, for their cooperation and for the Aureomycin used.

change from the special diluent to normal saline solution, as a vehicle, has apparently eliminated further serious consideration of anemia in relation to Aureomycin.

In the dosage used, we found that Aureomycin was non-toxic. Very few patients complained of pain at the site of injection and none had any systemic reaction. Its daily use over a three-month period of time did not produce any damage to any of the body systems; in particular, we observed no allergic skin reactions, no evidence of agranulocytosis, nor any changes in the peripheral or central nervous systems.

This use of Aureomycin is of great significance because it represents, for the first time, the high probability of a cure of a virus disease in a human being. Data bearing on this subject were presented in our original paper. The clinical results reported in this paper lend additional weight to this point. Our results were very gratifying and seemed to parallel those obtained by Wong and Cox³ who demonstrated that Aureomycin showed remarkable therapeutic value in mice, injected intracerebrally with the virus of lymphogranuloma venereum. In the light of the evidence at hand, we feel that if Aureomycin is used in the early stages of the disease and in proper dosage, it will cure certain, if not all strains of lymphogranuloma venereum infections. Further data bearing on proper dosage and period of administration is needed.

Material. The ten new cases of lymphogranuloma venereum infection consisted of four cases of buboes, two cases of acute proctitis and four cases of rectal stricture.

The diagnosis, in each instance, was made by the usual clinical criteria—and all patients gave positive Frei reactions.

Treatment. Eight patients received 20 mg. of Aureomycin daily, by the intramuscular route. Aureomycin was used orally in two cases in the following manner: a 300 mg. dose was given every four hours—i.e., using 5 mg. per kilo of body weight.

The results obtained in these ten cases demonstrated, again, the efficacy of the drug as a specific therapy for this disease.

Lymphogranuloma Case Reports. Brief abstracts of these ten cases are as follows:

BUBO CASES:

Case 1. A twenty-four-year-old male entered the hospital, with enlarged inguinal nodes, which had been present for two weeks. On the right side, enlarged inguinal nodes measured 5 cm. in diameter. On the left, the nodes measured 2.5 cm. in diameter. There was a urethral discharge which was negative for gonococci, on smear and culture.

The patient was given 300 mg. of Aureomycin, orally, four times a day, from May 8th to May 14th. At the end of the third day of treatment, the nodes on the right side had reduced to 4 cm. in diameter. By the end of the fifth day of treatment, all tenderness had disappeared and the nodes

on both sides were reduced to half their previous size. The urethral discharge had disappeared under treatment.

This patient then left the hospital, against advice, at the end of six days, and did not return to the follow-up clinic.

Case 2. A twenty-nine-year-old female entered the hospital, complaining of vaginal discharge and a swollen, tender inguinal mass of two and one-half weeks' duration. Physical examination revealed a matted, tender group of nodes in the right inguinal region which measured about 2 x 3.5 cm.

The patient was given 300 mg. of Aureomycin, orally, four times a day from April 28th to May 9th, 1948. The tenderness in the nodes disap-

TABLE 1
FOUR NEW CASES OF BUBOES

<i>Case No.</i>	<i>Age sex</i>	<i>Days on service</i>	<i>Lesion duration</i>	<i>Diagnosis</i>	<i>Dosage</i>	<i>Results</i>
1	24-M	7	2 weeks	Rt. bubo 2" diameter— Lt. bubo 1" diameter— Urethral discharge	300 mg. Q.I.D. orally Total 8,400 mg.	After 3 days, rt. 1½" diameter— After 5 days, rt. 1" diameter— lt. ½" diameter.
2	29-F	11	2½ weeks	Rt. bubo 2 x 3½ cm. Vaginal discharge	300 mg. Q.I.D. orally Total 14,400 mg.	After 1 day, node 2 x 3 cm.— After 72 hours, node 1½ cm. in diameter— After 8 days, node disappeared.
3	21-M	23	4 weeks	Lt. bubo 4 x 6 cm.	20 mg. I.M. O.D. Total 460 mg.	After 4 days, mass non-tender— 50% reduction in size
4	28-F	7	2½ weeks	Lt. bubo 5 cm. in diameter	20 mg. I.M. O.D. Total 140 mg.	After 3 days, mass non-tender— 50% reduction in size.

peared in twenty-four hours and in three days the node diameter was half of its previous size. By the eighth day the mass had disappeared.

The patient was discharged on the eighth day of hospitalization and did not return to the follow-up clinic.

The patient received a total of 14.4 grams of Aureomycin.

Case 3. A twenty-one-year-old male entered the hospital with a history of swollen nodes in the left groin of four weeks' duration. Physical examination revealed a tender, matted mass of left inguinal nodes, measuring 4 x 6 cm.

The patient was given a daily intramuscular injection of 20 mg. of Aureomycin from June 15th, 1948, to July 8th, 1948. At the end of the

fourth day of treatment, the tenderness had disappeared and the mass had shrunk to half of its previous size.

On July 17th, at the time of last examination, there was present a small area of induration 2.5 cm. in diameter, and the patient was symptom-free.

Case 4. A twenty-eight-year-old female entered the hospital, complaining of a tender mass in the left groin. Physical examination revealed a tender, matted mass of nodes in the left inguinal region 5 cm. in diameter.

The patient was given a daily injection of 20 mg. of Aureomycin from May 12th to May 19th, 1948. At the end of the third day of treatment, the mass decreased to one-half the original size and was non-tender.

After seven days of treatment the patient was so improved that she left the hospital, against advice. We have not seen her since.

(The bubo cases are tabulated in TABLE 1.)

PROCTITIS CASES:

Case 5. A thirty-four-year-old female entered the hospital on April 28th, 1948, complaining of a foul-smelling rectal discharge and rectal pain of two months' duration. Physical examination revealed a severe proctitis. On proctoscopic examination, the rectal mucosa was granular in appearance, and covered with yellowish plaques of pus. There was a thin, foul-smelling rectal discharge.

The patient was given daily injections of 20 mg. of Aureomycin from April 29th, 1948, to May 9th, 1948. Two days after the beginning of treatment, there was a marked reduction in the rectal discharge and pain. The discharge disappeared at the end of four days of treatment and, at this time, proctoscopic examination revealed normal rectal mucosa.

The patient left the hospital, greatly improved, but we have not seen her since.

Case 6. A forty-two-year-old female entered the hospital with a five-week history of rectal pain and discharge. Physical examination revealed

TABLE 2
TWO NEW CASES OF PROCTITIS

Case No.	Age sex	Days on service	Lesion duration	Diagnosis	Dosage	Results
5	39-F	19	2 months	Proctitis— with profuse rectal dis- charge	20 mg. I.M. Q.I.D. total 220 mg.	Marked reduction of discharge after 48 hours. Complete disap- pearance after 4 days. Proctoscopic examination showed normal rectal mucosa.
6	42-F	12	5 weeks	Proctitis— with rectal bleeding and pain	20 mg. I. M. Q.I.D. total 220 mg.	Rectal bleeding and pain disap- peared after 7 days.

a protuse, thin, seropurulent, foul-smelling rectal discharge. She had a soft, tender, bleeding stricture about one and one-half inches from the anal verge.

She was given a daily intramuscular injection of 20 mg. of Aureomycin from June 26th, 1948, to July 6th, 1948. After seventy-two hours, there was a marked decrease in rectal pain and discharge. On the eighth day after the onset of treatment, rectal bleeding and pain disappeared. The soft stricture was not to be felt, and proctoscopic examination was negative.

In this instance, the stricture was obviously due to mucosal and submucosal edema, which subsided under treatment. She was discharged, apparently cured.

(See TABLE 2 for these cases.)

RECTAL STRICTURE CASES:

Case 7. A forty-eight-year-old female was admitted to the hospital on April 20th, 1948, complaining of constipation. On examination, a hard, fibrous rectal stricture, which did not admit a fingertip, was palpated above the anus.

She was given 20 mg. of Aureomycin, intramuscular, for a period of sixteen days. After 320 mg. had been given, there was no change in the stricture, although the patient felt better. The constipation persisted and a colostomy was performed.

She was discharged from the hospital on June 12th, 1948, improved.

Case 8. A thirty-eight-year-old female complained of constipation and of having pencil-sized stools. She was admitted to the hospital on May 17th, 1948. On examination, she had a rectal stricture which did not admit the tip of a finger. She was observed for thirty-five days in our hospital.

She received 20 mg. of Aureomycin each day for twenty-eight days—a total dose of 560 mg. The stricture did not improve. However, the secondary infection disappeared under therapy, and on the twenty-fifth day of hospitalization she was colostomized.

Case 9. A sixty-two-year-old female entered the hospital on May 19th, 1947, complaining of severe constipation, which had been growing worse, for six years. Examination showed a rectal stricture 3 cm. above the ano-cutaneous junction. This did not admit a fingertip.

She was given Aureomycin in 20 mg. doses daily, by intramuscular injection, until she had received a total dosage of 140 mg. Under treatment, the diameter of the stool doubled and she reported that bowel movements were "easier than in years." She was subjectively improved and left the hospital, against advice, seven days later.

Case 10. A thirty-one-year-old female entered the hospital, complaining of the passage of stools per vagina, for six weeks. She gave a history of having had rectal trouble for ten years, and of a suprapubic hysterectomy, performed at another hospital, three years ago. On rectal examina-

tion, she showed a stricture that had closed down tightly and did not admit a fingertip. It was hard on palpation. Vaginal examination showed a recto-vaginal fistula, which measured about 1.5 cm. in diameter.

She received 20 mg. of the drug in daily doses, beginning June 20, 1948, until a total dose of 400 mg. had been given. There was no change in the stricture, and, as was to be expected, there was no improvement in the fistula. Obviously a colostomy was indicated in this case.

(See TABLE 3 for these four rectal stricture cases.)

TABLE 3
FOUR NEW CASES OF RECTAL STRICTURES

Case No.	Age sex	Days on service	Lesion duration	Diagnosis	Dosage	Results
7	48-F	53	Many years	Rectal stricture does not admit fingertip	20 mg. I. M. Total 320 mg.	Colostomy performed. Secondary infection improved. No improvement of fibrotic stricture.
8	38-F	35	5 years	Rectal stricture does not admit fingertip	20 mg. I. M. Total 560 mg.	Colostomy performed. Secondary infection improved. No improvement of fibrotic stricture.
9	62-F	36	10 years	Rectal stricture does not admit fingertip	20 mg. I. M. Total 140 mg.	Diameter of stools doubled in size. Home against advice. No change in stricture.
10	31-F	20	10 years	Rectal stricture Recto-vaginal fistula	20 mg. I. M. Total 300 mg.	Secondary infection improved. No change in stricture.

The results obtained in the treatment of these ten new cases, whether by the intramuscular or oral route, were similar, in every regard, to the results obtained in our previous series of twenty-five cases. The acute manifestations of lymphogranuloma venereum, namely the buboes and proctitis cases, were invariably cured. In the cases of rectal stricture, the secondary infections were cured and no extension of the lymphogranulomatous process was observed.

It is needless to point out that no antibiotic can cure a thick, fibrous rectal stricture. In no instance should surgery be withheld when indicated, whether it be colostomy or resection of the stricture. Severe mechanical conditions are the only indication for surgery. It is clear to us that Aureomycin should be used before surgery in all cases of benign fibrous stenosis of the rectum.

FOLLOW-UP RESULTS IN FOURTEEN OF TWENTY-FIVE PREVIOUSLY REPORTED CASES

We were able to follow up fourteen of the first twenty-five cases treated which had been recorded in our first, as yet unpublished, paper on this subject. We made every effort to get the other cases to return to our follow-up clinic, but they did not do so. The follow-up period varied from two weeks to four months after discharge from the hospital and cessation of the administration of the drug. The follow-up material is comprised of two cases of buboes, one case of proctitis, and eleven cases of rectal stricture.

Two cases buboes were followed in the Out Patient Clinic. In one case, the node had been removed for study and when the patient was examined four months later, this operative wound had healed, *per primum*, and he had no complaints. The other bubo case, when examined two months later, was cured and the patient was in excellent condition. We have reason to believe that the other bubo cases had improved so much that they did not return to the follow-up clinic.

Only one proctitis case returned to the follow-up clinic. He showed a recurrence of symptoms, as evidenced by rectal discharge. In view of the fact that he is an admitted homosexual, who has been the passive partner in the practice of sodomy, it is not clear as to whether this is a true recurrence or a new infection.

Eleven cases of rectal stricture have been followed for period of time varying from two weeks to sixteen weeks. The average period of time since the discontinuance of the use of the drug was about eight weeks.

Nine patients have maintained the good results obtained under treatment. They had no rectal pain, discharge or bleeding and, where the stool diameter had increased under treatment, this increase remained. In several cases where, on initial examination, the stricture did not admit the tip of the examining finger, it did admit the examining finger at discharge from the hospital and throughout the follow-up period. All patients stated that they felt much better and considered the result good.

In one instance, there was an apparent relapse. This patient was admitted with a very soft stricture. She remained in the hospital three months and received during this time, 320 mg. of Aureomycin intramuscularly. On discharge, the stricture could not be palpated and the patient was apparently cured. One month later she was seen in our follow-up clinic, complaining of rectal pain, bleeding, and discharge. On rectal examination, no stricture was found, but proctoscopic examination showed an ulceration of the mucosa. She refused further treatment. It is therefore impossible for us to say whether this instance was a relapse or a reinfection. Certainly the stricture, which was felt on admission, was due to mucosal and submucosal inflammation, and not to scar tissue.

In one other instance, there was a recurrence of rectal discharge observed in a patient who returned to the follow-up clinic. This patient had

had a colostomy and, while in the hospital, her rectal discharge disappeared after she had received 210 mg. of Aureomycin intramuscularly. During her hospital stay she gained twenty pounds in weight. Her discharge recurred, subsequent to leaving the hospital, but the patient refused further treatment.

In general, it may be stated that most of our rectal stricture cases gained weight, even without the benefit of colostomy, and this gain in weight was maintained during the follow-up period.

In all of these cases, Aureomycin not only cleared up the secondary infection, but not in any single case was there seen an extension of the disease process clinically, in so far as we could determine. It is still too early to draw final conclusions as to the effect of Aureomycin on the extension of the disease process.

Granuloma Inguinale

In view of the unusual therapeutic range of Aureomycin and its low toxicity, it seemed desirable to use this antibiotic in patients suffering from granuloma inguinale. Other types of therapy which have been used in this disease—chemotherapy, tartar emetic, Fuadin (Winthrop), Anthomaline (Merck), radiation therapy and podophyllin in olive oil—were unsatisfactory because of the frequency of relapses and the chemoresistance of the chronic cases. We felt that Aureomycin might be effective but less toxic than streptomycin in this condition.

Three cases of granuloma inguinale have been treated by us with Aureomycin with eminently satisfactory results. This is the first use of Aureomycin to our knowledge for the treatment of this disease in human beings.

The abstracts of the case histories are as follows:

Case 1. The patient is a thirty-nine-year-old white seaman who entered the hospital on April 1st, 1948, complaining of an ulcer of the foreskin of the penis. He gave the following history:

The ulcer had appeared four months previously in January of 1948. At first he treated it with sulfathiazole ointment, soap and iodine. He saw a doctor on January 12th, 1948, who treated him with one of the sulfa drugs for six days. At that time, the Frei and Wassermann tests were repeatedly negative. His condition did not improve and the patient, therefore, entered a hospital in Galveston, Texas. There, blood and spinal fluid tests were negative. He was treated with penicillin therapy, but the ulcer did not respond.

On March 8th, 1948, he entered a Staten Island hospital. A report from this institution stated that darkfield examinations for *Treponema pallidum* were negative. Blood and spinal fluid tests for syphilis, the Frei test for lymphogranuloma venereum, and smears for the Ducrey bacillus were all negative. The patient refused biopsy of the lesion. At this time, he gave the additional history of a gonorrheal infection and a penile lesion for a period of three months in 1927, for which he had received arsenic

and mercury treatment. Because it was considered that this treatment was probably inadequate, the patient was given 40,000 units of penicillin, every two hours, for eighty-five doses. He left the hospital, against advice, because the ulcer did not improve.

Examination at Harlem Hospital showed no abnormal findings except a tender ulcer (see FIGURE 1), which measured 4 x 6 cm. Biopsy was done on April 7th, 1948, and Donovan bodies were found.

On April 10th, 1948, 20 mg. of Aureomycin was administered by the intramuscular route and given daily thereafter. Two days after treatment was initiated, there was marked improvement, the tenderness had disappeared, and the patient felt better. On the third day of treatment, the base of the ulcer was dry and it was greatly reduced in size. Six days later there was no further discharge from the ulcer and the lesion was one-quarter of its previous size.

On April 24th, 1948, the ulcer measured $1\frac{1}{2}$ x 1 cm. and the dose of Aureomycin was increased to 20 mg. twice a day. On May 8th, 1948, the base of the lesion was indurated, although still smaller in size. It was non-tender and there was no discharge. The patient had received, by this time, 1.06 grams of Aureomycin, with no toxic reactions.

On June 4th, 1948, after 2.02 grams of Aureomycin had been administered, the ulcer was completely healed.

The patient was discharged on June 7th, 1948 (see FIGURE 2). He has been followed since, but there is no sign of recurrence.

Case 2. The patient, a thirty-two-year-old female, was admitted with a history of known granuloma inguinale for the past seven years. At its inception, she had experienced itching around the vagina and, two days later, an eruption appeared, which spread rapidly to the perineal and perirectal regions.

In 1942, a diagnosis of granuloma inguinale was made in our Out Patient Department. The lesion had healed under Fuadin treatment, but two weeks after cessation of treatment, it recurred. She has had numerous courses of Fuadin since, only to have the pathology recur at the conclusion of treatment.

Three months before the present admission, the patient developed an eruption in the skin of the pudendal and perineal region. She could not sit down because of the pain, so therefore she remained in bed or stood up most of the time.

On admission there was profuse, foul-smelling discharge from these granulating and ulcerating areas. In fact, the stench was so foul that, when her dress was raised for examination, the odor permeated a large

FIGURES 1-4 (see opposite page)

FIGURE 1. (Case 1) Ulcer of penis due to granuloma inguinale—before treatment.

FIGURE 2. (Case 1) Healed ulcer of penis after treatment.

FIGURE 3. (Case 2) Granuloma inguinale—Healed ulceration of perineum and vulva, posterior view—after treatment.

FIGURE 4. (Case 3) Granuloma Inguinale—View of healed ulcer of vulva—after treatment.



3



room. Laboratory studies revealed a negative Frei test and a positive Kahn. Other laboratory findings were within normal limits. Biopsy, on this admission, showed the presence of Donovan bodies in a giemsa stain.

On May 25th, 1948, the patient was started on oral Aureomycin—300 mg. every four hours. After four days she had no toxic reaction. At the end of six days of treatment the purulent discharge had ceased, and she felt subjectively better, being free of pain. The lesion was reduced about ten per cent in size. After three weeks of treatment, the lesion was seventy-five per cent healed.

The patient received a total dosage of 75.6 grams and was discharged completely healed, and with a sense of well-being (see FIGURE 3).

Case 3. A fifty-three-year-old female was admitted to Harlem Hospital with a history of perirectal swelling and discharge of ten years' duration. She had had two previous operations for removal of granulation tissue and a perianal fistula. Two years prior to the present admission to the hospital, the left labia became so swollen and tender that she could not walk, except with extreme pain. There was a foul-smelling discharge present. Examination was negative, except for pathological findings confined to the genitalia. The left labia were enlarged to three times their normal size and exuded a foul-smelling discharge. There was a thick, exquisitely tender, perirectal mass. Laboratory studies revealed a positive Frei test and the Wassermann reaction was negative. Other laboratory findings showed normal values. Biopsy was done and revealed the presence of Donovan bodies.

The patient was given 20 mg. of Aureomycin daily for twenty-eight days, receiving a total dosage of 560 mg. After eight days of treatment, the patient was free of pain and there was a moderate reduction in the edema and in the foul discharge (see FIGURE 4).

At the time the patient left the hospital, the discharge had completely disappeared and her vulva had returned to normal.

CLINICAL RESULTS:

The rapid and satisfactory response of these three proven cases of granuloma inguinale to Aureomycin seemed to indicate that it has a definite place in the treatment of this condition. Our follow-up has not been sufficiently long for us to comment upon the possibility of permanence of the cure. The good results, thus far obtained, demonstrate the necessity of further study of Aureomycin, particularly because it is non-toxic and can be taken orally (TABLE 4).

Oral Aureomycin. We have given Aureomycin, by the oral route, in two cases of lymphogranulomatous buboes and in one case of granuloma inguinale. The results were satisfactory and comparable to those obtained by intramuscular injections. The dosage was computed at 5 mg. per kilo of body weight. This dose was given every four hours. One patient received 8.4 grams and one other received 14.4 grams, while a third received 75.6 grams. No toxic reactions of any type were observed.

TABLE 4

THREE CASES OF GRANULOMA INGUINALE TREATED WITH AUREOMYCIN

Case No.	Age	Days since onset	Lesion duration	Diagnosis	Dosage	Results
1	39-M	38	4 months	Granuloma inguinale, penile ulcer	20 mg. I.M. Then 40 mg. I.M. Total 2,020 mg	Cured
2	32-F	43	7 years	Granuloma inguinale, vulvo-perineal	300 mg. Q.4.H. Total 75,600 mg.	Ulceration healed
3	53-F	28	10 years	Granuloma inguinale vulvo-perineal	20 mg. I.M. Total 560 mg	Ulceration healed Edema of vulva reduced

TABLE 5

CASE 2—ORAL AUREOMYCIN 300 MG.

		Tube dilution						Activity (micrograms ml.)
		1	2	4	8	16	32	
Before drug	Undilute	++	++	++	++	++	++	<0.05
1 hour after 300 mg. orally	Undilute	—	—	—	—	—	+	
	1—10	—	—	+	++	++	++	1.0
2 hours	Undilute	—	—	—	—	—	—	
	1—10	—	—	—	+	++	++	2.0
3 hours	Undilute	—	—	—	—	—	+	
	1—10	—	—	+	++	++	++	1.0
4 hours	Undilute	—	—	—	—	—	+	
	1—10	—	—	+	++	++	++	1.0
300 mg. orally 1 hour after second dose	Undilute	—	—	—	—	—	+	
	1—10	—	—	+	+	++	++	1.0
2 hours after second dose	Undilute	—	—	—	—	—	—	
	1—10	—	—	—	+	++	++	2.0

Blood Levels. A sampling of the blood levels of Aureomycin was done on two patients, one on oral therapy and the other on intramuscular therapy. The method used in carrying out this work on blood levels was developed by Dornbush⁴ and was done in his laboratory.

In micrograms of activity, per milliliter, the highest blood level, namely two micrograms per milliliter, was obtained two hours after oral administration of 300 mg. of Aureomycin, as shown in TABLE 5.

In the other patient, who received 20 mg. intramuscularly, the highest concentration of the drug, namely one microgram per milliliter, was noted five hours after injection of the drug, as shown in TABLE 6.

TABLE 6

CASE 3 -20 MG. by I.M.

		Tube dilution						Activity (micrograms/ ml.)
		1	2	4	8	16	32	
Before dose	Undilute							
24 hours after last dose	Undilute	++	++	++	++	++	++	<0.05
1 hour	3—10	—	—	+	++	++	++	0.33
2 hours	Undilute	—	—	—	—	+	++	
	1—10	—	+	++	++	++	++	0.5
3 hours	Undilute	—	—	—	—	+	++	
	1—10	—	+	++	++	++	++	0.5
4 hours	Undilute	—	—	—	—	+	++	
	1—10	—	+	++	++	++	++	0.5
5 hours	Undilute	—	—	—	—	—	+	
	1—10	—	—	+	++	++	++	1.0
6 hours	Undilute	—	—	—	—	—	+	
	1—10	—	—	+	++	++	++	1.0

No true comparison can be made because of the relatively massive dose given orally in contrast to the small dose given intramuscularly.

Summary

1. Ten new cases of lymphogranuloma venereum treated with Aureomycin are reported, in detail. Together with the twenty-five cases reported in a preliminary paper, this makes a total of thirty-five cases. The results were excellent and we believe that this antibiotic is a superior, specific form of therapy for the lymphogranuloma venereum virus infection. This antibiotic is also very effective against the secondary bacterial invaders present in this disease.

2. Fourteen of the original twenty-five cases reported in our, as yet unpublished, paper have been followed for periods of time varying from two to sixteen weeks after discharge from the hospital. These cases show that the curative effects of Aureomycin persist after treatment has been stopped.

3. Ulcerative lesions in three patients with proved granuloma inguinale were healed by the use of Aureomycin. Its lack of toxicity, in the dosage used, and also the fact that it can be taken orally, demand further study of this antibiotic in this disease.

Conclusions

Aureomycin is the treatment of choice in all cases of lymphogranuloma venereum infection. If mechanical conditions demand, surgery, in conjunction with Aureomycin, is indicated.

Aureomycin will heal ulcerations produced by the causative agent of granuloma inguinale. Further extensive clinical research in the use of Aureomycin in the treatment of granuloma inguinale is warranted.

Bibliography

1. WRIGHT, L. T., M. SANDERS, M. A. LOGAN, A. PRIGOT, & L. M. HILL. Aureomycin: A New Antibiotic with Virucidal Properties. A Preliminary Report on The Successful Therapy in 25 Cases of Lymphogranuloma Venereum. (In Press.)
2. WRIGHT, L. T., W. A. FREEMAN, & J. V. BOLDEN. 1946. Lymphogranulomatous strictures of the rectum. A résumé of 476 cases. Arch. Surg. 53: 499-544.
3. WONG, S. C., & H. COX. 1948. Ann. N. Y. Acad. Sci. 51(2): 290-305.
4. DORNBI'GH, A., & E. J. PELCAK. 1948. Ann. N. Y. Acad. Sci. 51(2): 218-220.

TREATMENT OF Q FEVER IN MAN WITH AUREOMYCIN*

By E. H. LENNETTE, G. MEIKLEJOHN, AND H. M. THELEN

*California State Department of Public Health, Viral and Rickettsial Disease Laboratory,
Berkeley, and Division of Medicine, University of California Medical School,
San Francisco, California*

AN opportunity to evaluate aureomycin therapy, found by Wong and Cox¹ to be effective in experimental Q fever, was afforded by the occurrence of human cases of Q fever in southern² and northern California.³ In the northern part of the state, where this study was conducted, cases of the disease were first recognized in considerable numbers early in 1948³ and to the time of writing more than 125 cases have occurred.

Studies with aureomycin were undertaken in the middle of May, 1948, and thus far, 23 patients have been treated. The preliminary results of this study form the basis of the present report.

Plan of Study

It was considered impractical, for a number of reasons, to treat alternate patients with aureomycin. The relatively small number of patients available at any one time, the large geographic area over which they were scattered, and the lack of knowledge of desirable doses of the drug by the parenteral and oral routes were the determining factors in this decision. It was believed that evidence for the effectiveness, or lack of effectiveness, of aureomycin might be obtained more rapidly by selecting for treatment a group of patients whose course was such that a severe, prolonged illness appeared likely. In the event that the illness terminated in each instance within a relatively short interval, a case for the effectiveness of the drug in Q fever could be established. Such a procedure obviously necessitated a comparative study of numerous untreated individuals in order to obtain as complete a spectrum as possible of the clinical picture. Through the cooperation of physicians in three different areas where cases were occurring, pertinent clinical information was eventually accumulated on a total of 90 patients with a serological diagnosis of Q fever, and when collated gave some conception of the type of illness being encountered in different age groups in the 3 localities. Patients to be treated were chosen in the older age

* The authors are indebted to Drs. O. C. Railsback and A. M. Clark of the Woodland Clinic, Woodland, Drs. C. R. McKinney and T. Y. Cooper of Davis, and Drs. J. L. Faulkner and R. G. Frev of Red Bluff for their generous assistance and cooperation in this study. They are also indebted to Dr. C. L. Collins for making available to them the facilities of the Permanente Foundation Hospital at Vallejo.

The authors also wish to thank Mrs. Beatrice England, Mrs. Jeane Harris, and Mrs. Florence Jensen for technical assistance in the serological studies and Mr. Oscar Brunetti and Miss Margery Maggs for technical assistance in the animal work.

The aureomycin was supplied by Dr. Herald R. Cox.

groups (*vide infra*) in each of the 3 areas, and the clinical diagnosis in all instances was checked by complement-fixation tests for Q fever.

The duration of fever was selected as the best objective criterion for the evaluation of the therapeutic effect of the drug, and in this connection several points on the febrile course in untreated patients should be mentioned. During the early phase of illness, the temperature was usually high, often exceeding 104° F. In certain instances, the fever terminated within a few days, while in others, the initial high febrile phase was followed by a prolonged period of low grade fever. These differences were, in major part, associated with age, as will be brought out below.

The use of the duration of fever as the sole criterion of the patient's response to therapy has obvious limitations, since the marked symptomatic improvement which may occur before the fever has completely subsided is not reflected in the temperature graph.

Course in Untreated Patients

TABLE 1 shows the duration, by age groups, of the febrile course in the 90 patients comprising the untreated group. With due allowance for the fact that the individuals with a severe illness were more likely to seek medical aid, and thus weight the group in favor of the more severe course, it is still apparent that Q fever as seen in central and northern California has frequently been a disease with a protracted course, especially in older individuals. The short duration of the disease in the younger age groups is similar to that reported in military groups, and contrasts sharply with the course in older individuals.

TABLE 1

DURATION OF FEBRILE COURSE IN UNTREATED PATIENTS, ACCORDING TO AGE

Age group	Number of febrile days*			Total number patients
	7 or less	8-14	15 or more	
25 or less	38	10	3	51
26-35	11	6	5	22
36 or more	3	2	12	17
Total	52	18	20	90

* Febrile day defined as one on which one or more oral temperature readings were 99.3° F. or higher.

It was early apparent that in young individuals, *i.e.*, those under 25 years of age, the course of the disease was subject to such extreme variation, and was so frequently short, that comparative studies of treated with untreated patients would present great difficulties. Among individuals 26 or older, on the other hand, such a large proportion had an illness of considerable severity and long duration, that this age group appeared to offer the best opportunity for comparative studies, and consequently patients to be treated were selected, with certain exceptions noted below, from this older age group.

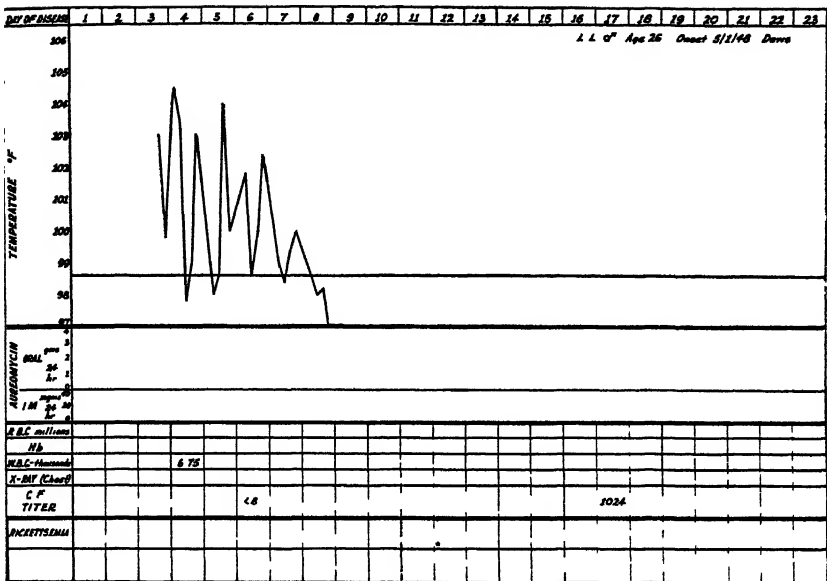


FIGURE 1. A mild case of Q fever (untreated).

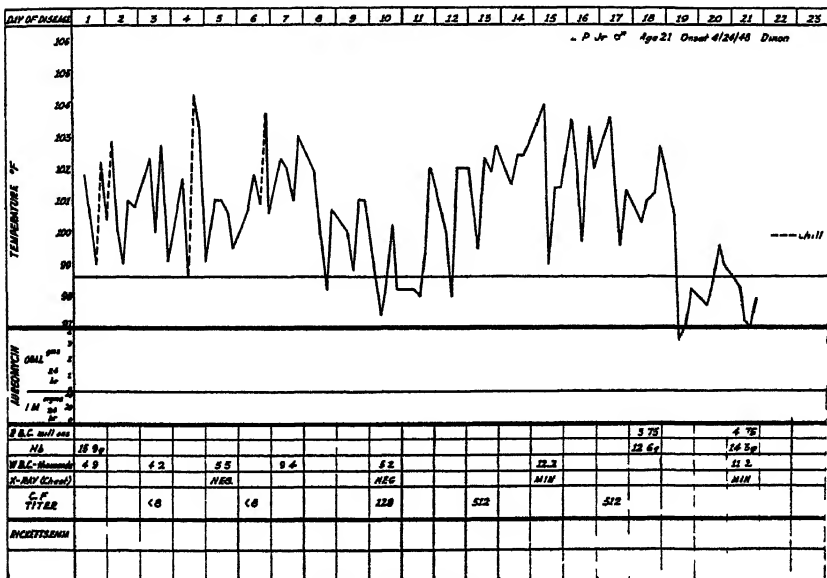


FIGURE 2. A moderately severe case of Q fever (untreated).

FIGURES 1 and 2 present data on a mild and on a moderately severe case, respectively, of Q fever which did not receive aureomycin. Like most of the patients in the "untreated" group, they received penicillin or a sulfonamide drug.

Course in Treated Patients

The patients to whom aureomycin was given were chosen in the light of experience with the untreated group in an effort to pick those individuals most likely to have a long and stormy course. As mentioned above, these individuals were primarily in the older age groups, and since the disease was being studied in 3 different areas, the treated patients were compared with untreated persons of the same age group in the same area in order to obviate, insofar as possible, unknown factors associated with geography, exposure to infection, mode of transmission, differences in infecting strains, etc.

With the exception of two younger patients, selected because they were unusually ill, all were over 25 years of age and the majority were over 30 years of age. All but one (J. R., TABLE 2) had been ill for 4 or more days, and all were becoming increasingly ill rather than showing improvement. All had had one or more true rigors and, except for 1 individual (R. S., TABLE 2), had had oral temperature readings of 104° F. or more.

TABLE 2

RESULTS IN PATIENTS TREATED DURING ACUTE PHASE WITH ORAL AUREOMYCIN

No.	Patient	Age	Max. temp.	Chills	Febrile days		
					Before treat- ment	After treat- ment	Total
1.	J. R.	22	104.4	+	2	2	4
2.	H. T.	22	104.4	+	4	3	7
3.	G. N.	28	104	+	4	3	7
4.	L. T.	48	104	+	4	3	7
5.	J. P.	27	105	+	5	0	5
6.	G. S.	63	104.4	+	6	4*	10
7.	C. G.	48	104	+	7	5	12
8.	F. M.	65	104	+	8	2	10
9.	R. J.	39	104	+	8	3*	11
10.	R. S.	28	103	+	9	3	12

* Relapse following cessation of therapy, see text

Of the 23 patients treated to date, 18 received aureomycin within 12 days of the onset of illness. In this group of 18, 2 received intramuscular therapy only, 2 intramuscular, followed by oral therapy and 14 oral therapy only. The remaining 5 patients, who had been febrile for from 23 to 77 days after onset, were treated orally.

In the following discussion, data are presented only on those patients, 14 in number, who were treated early in the disease and on whom etiological studies confirming the clinical diagnosis have been completed.

All of these have been followed for more than two weeks after cessation of therapy.

Intramuscular Therapy.—The first 4 patients were treated with 40 mg. of aureomycin hydrochloride daily given in two injections of 20 mg. each at 12-hour intervals. Two patients (R. B. and J. C.) showed prompt symptomatic improvement with fall of temperature to normal within 72 hours. A third (A. B.) also improved, but continued to show daily low elevations of temperature until placed on oral therapy. The fourth patient (D. D.) showed no improvement during 4 days on this regime; he was then put on oral therapy and his temperature fell to normal within 48 hours. Severe pain at the site of injection, even when phosphate buffer (pH 7.2) was used as the diluent, was the cause for discontinuing parenteral therapy in the latter two patients.

The results obtained with these small intramuscular injections are not considered convincing.

Oral Therapy.—Eight patients over 25 years of age, with a spread from 28 to 65 years, and 2 unusually ill patients under 25 were treated with aureomycin by the oral route.

These individuals received during the first 24 hours a dose of 3.2 or 4.0 g. of the drug, and were then maintained on 1.6 or 2.0 g. per day for 4 or more days. The smallest total dose of drug administered was 8.0 g., the largest 27.5 g.

In every instance, symptomatic improvement, manifested best by the return of appetite, was noted within 48 hours, and a considerable decline of temperature occurred within 48 to 72 hours after therapy was commenced. The temperatures of 8 patients fell to within normal limits (99° F. or below) during the first 3 days of therapy and the temperatures of the other 2 became normal after 4 and 5 days of therapy. The data on these 10 patients are summarized in TABLE 2.

The drug was discontinued in some instances within 24 hours after the temperature became normal. In 8 of the 10 patients presented in TABLE 2, convalescence was uneventful. In the remaining 2 (G. S. and R. J.) who had received only 10.0 g. of the drug over a 4 day period, fever recurred 2 days after the drug had been stopped. Therapy was re-instituted and the temperature returned to normal in much the same pattern as had been observed during the first episode.

The course of the illness in 5 patients is illustrated in FIGURES 3-7. and a brief description of each of these patients follows.

Case 3 (FIGURE 3). G. N., male, 28, a student, became ill on May 24, 1948, and was admitted to the infirmary, that evening, with complaints of headache, malaise, and feverishness. During the next four days his temperature fluctuated repeatedly from normal to peaks of 104° and he had repeated shaking chills. Physical examination, at this time, did not reveal any abnormal findings in chest or abdomen. The white blood cell count was 4600.

Oral aureomycin therapy was begun on May 27, 3.6 g. being given



during the first 24 hours and 1.8 g. daily thereafter for 4 days, a total of 10.8 g. On May 28, the patient again had a chill and his temperature rose to 104° F. At this time, he appeared acutely ill, the liver was palpable 1 fingerbreadth below the costal margin and was slightly tender, and the spleen was felt 1 fingerbreadth below the costal margin. The following day, he felt considerably better. His temperature returned to within normal limits on May 31. For several days following discharge on June 2, the patient complained of weakness and sweating, but his temperature remained normal.

Etiological data: Inoculation into guinea pigs of blood taken on May 26 showed the presence of *Coxiella burneti*. In addition, complement-fixing antibodies to *Coxiella burneti* appeared in the patient's blood. The titer rising from <1:8 on May 25 to 1:512 on June 15.

Case 4 (FIGURE 4). L. T., male, 47, a plumber, became ill on June 2, 1948, with complaints of malaise, headache, feverishness, chilliness, pain in the eyeballs, and generalized aching. He was admitted to the hospital on June 25 following a shaking chill which lasted from 10 a.m. to noon. On physical examination at this time, the patient appeared fatigued, the skin was wet and hot. No abnormal findings were discovered on examination of the heart, lungs, and abdomen. The white blood cell count was 6,000.

Oral aureomycin therapy was begun on June 5, 4.0 g. being given during the first 24 hours and 2.0 g. per 24 hours for the succeeding 3½ days. On June 7, the patient felt considerably improved and his appetite was returning. The temperature became normal on the morning of June 8 and remained so thereafter.

Etiological data: Guinea pigs inoculated with a blood specimen taken on June 5 subsequently developed complement-fixing antibodies to *Coxiella burneti*.

Case 6 (FIGURE 5). G. S., male, 63, a slaughter-house worker, noted malaise and feverishness on May 29, 1948. During the following 4 days, he experienced repeated rigors, his temperature rose to 104° F. each evening and he complained of severe frontal headache, a non-productive cough, pains in the loins, and nausea. He was hospitalized on June 2. Physical examination at that time revealed an exceptionally robust man, who did not appear severely ill, but who was uncomfortable and perspiring profusely. Harsh breath sounds were noted over the left base posteriorly. The spleen was felt 1 fingerbreadth below the costal margin, and was firm and not tender. The white blood cell count was 5600. Roentgenograms of the chest failed to show any definite area of increased density.

Oral aureomycin therapy was begun on June 2, 4.0 g. being given during the first 24 hours and 2.0 g. daily thereafter for 3 days. During this period, marked symptomatic improvement was noted, appetite returned, and the patient's temperature returned to normal. His only complaints were a sore throat, soreness and a numb sensation referred

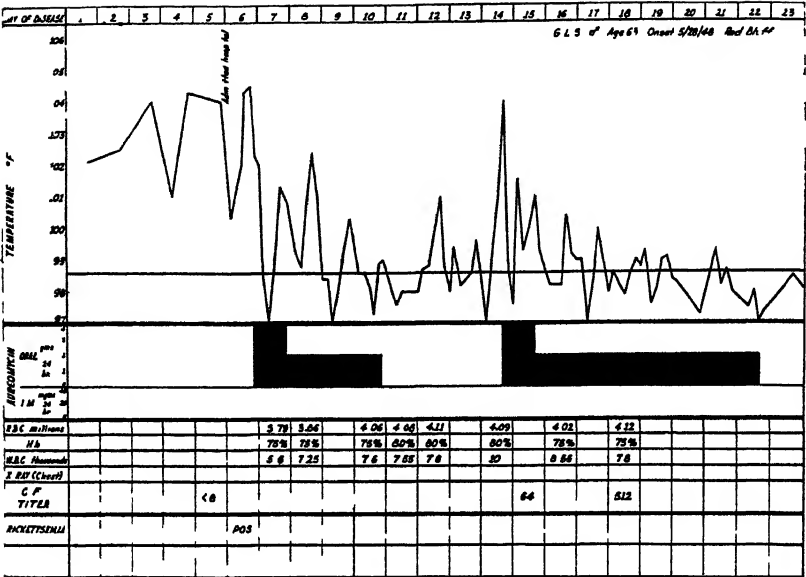


FIGURE 5. Relapse following cessation of therapy, and course after resumption of aureomycin treatment (Case 61).

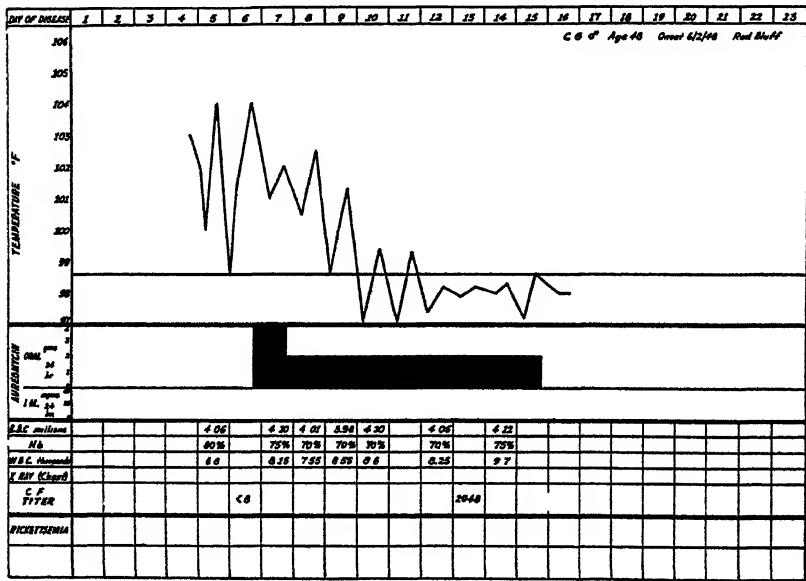


FIGURE 6. Course in patient treated with aureomycin (Case 7).

to the tongue, and a burning sensation on the surface of the scrotum. The tongue appeared smooth on the edges and was fiery red, as was the pharynx, and the scrotum presented a red, beefy appearance.

Following cessation of therapy, the patient's temperature remained normal for one day, but thereafter he began to have spikes of fever and on June 10 his temperature again reached 104° F. Oral aureomycin was resumed on the same schedule as before, and his temperature fell to normal during the succeeding three days. The drug was continued until June 18, a total of 17.5 gm. being given during this second course. The patient again noted some soreness and numbness of his tongue, but this tended to lessen during the period of therapy. His temperature remained normal following discharge from the hospital, and convalescence was uneventful.

Etiological data: Complement fixation tests for Q fever gave the following results: Titer, June 1, <1:8; June 11, 1:64; June 14, 1:512.

Case 7 (FIGURE 6). C. G., male, 48, a slaughter-house worker, became ill on June 2, 1948. He was admitted to the hospital on June 5 with complaints of malaise, feverishness, chills, headache, soreness of the eyeballs, generalized aching, anorexia and nausea. He coughed occasionally, raising small amounts of sputum. He had had severe shaking chills on 3 occasions. The white blood cell count was 6800. He was given penicillin, 30,000 units intramuscularly, every 3 hours for 3 days, but failed to show improvement.

When examined on June 8, his temperature was 102° F., pulse 76, respirations 80. He was a well-developed man who appeared moderately ill and uncomfortable. Fine *râles* were heard in the right posterior axillary region and the breath sounds were harsh in that area. The liver edge was felt 3 fingerbreadths below the costal margin and the liver was tender. The spleen edge was felt 1 fingerbreadth below the costal margin and was tender. Other physical findings were not abnormal.

Oral aureomycin was started on June 8, 4.0 g. being given during the first 24 hours and 2.0 g. each 24 hours thereafter, during the following 5 days, a total of 14.0 g. On June 10, the patient was more comfortable and his appetite had returned. The spleen was not felt, but the liver remained enlarged and tender. He was afebrile from June 13 on. Convalescence was uneventful.

Etiological data: On June 7, the patient's serum had a complement-fixing titer of <1:8, and on June 15, a titer of 1:2048.

Case 8 (FIGURE 7). F. M., male, 65, a railway shipping clerk, became ill on June 2, 1948, complaining of headache, pains in the loins, feverishness, repeated chills, nausea, and occasional vomiting. On June 9, following a shaking chill which lasted for two hours, his temperature rose to 104° F. On admission to the hospital at that time, he appeared moderately ill and fatigued. Physical examination failed to reveal any significant findings. White blood cell count was 5000.

He was placed on oral aureomycin therapy, 4.0 g. being given during

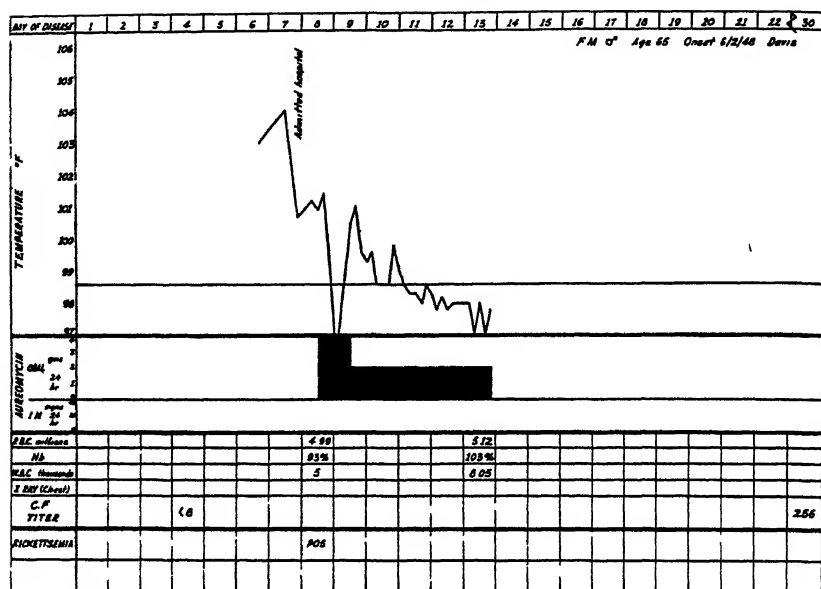


FIGURE 7. Course in patient treated with aureomycin (Case 8).

the first 24 hours and 2.0 g. daily thereafter during the following 4 days. The total dose of drug given was 12.0 g. Considerable symptomatic improvement was noted on June 10 and his temperature returned to normal on June 12. Convalescence was uneventful.

Etiological data: The Q fever complement fixation titer on June 5 was <1:8 and July 1 was 1.256.

Chronic Cases. Oral therapy has been used in 5 patients, mentioned above, who ran febrile courses of from 23 to 77 days. Four of these appeared to respond promptly. The fifth, a man of 51 years of age, who had been ill for 27 days when therapy was commenced, had failed to show, at the time of writing, appreciable response even with very large doses.

Toxicity

Mild symptoms referable to the gastro-intestinal tract were recorded in four cases. The chief complaints were a sense of fullness in the epigastrium with nausea and, occasionally, vomiting. Since all of these patients had had similar complaints prior to the beginning of therapy, the role of the drug in producing them is difficult to assess. One patient complained of pruritus and soreness of the scrotum and of soreness of the mouth. Both organs appeared inflamed. A second patient had similar complaints referable to the scrotum and also developed a small number of pruritic papular lesions over the shoulders. In both instances the drug was continued and symptoms did not become increasingly troublesome.

No evidence of hematological changes, of changes in urinary sediment, or of drug fever was observed

Discussion

The need for an effective therapeutic agent for Q fever has become a very real one in California. It is still premature, however, to state categorically that aureomycin will meet this need, since the number of cases which it has been possible to treat during a 2-month period under conditions permitting adequate study is still relatively small, and larger numbers are needed to establish the effectiveness of the drug in a disease which varies considerably in severity and duration, and apparently is rarely fatal.

At the present time, the drug appears to have had a definite therapeutic effect. First, in a number of exceptionally ill patients, symptomatic improvement, frequently striking, has been observed in virtually all instances soon after therapy was begun. Secondly, the duration of fever has been, with few exceptions, short, and the number of patients treated is sufficiently large to make it seem unlikely that this drop in temperature has been due to chance.

It should be noted that there have been some obvious differences in the rapidity with which patients have responded to treatment. It is not yet clear what factors are responsible for these differences. Under consideration are such factors as the stage of the disease at which therapy was begun and the failure of the patient to obtain sufficiently high blood levels of the drug. It may be more than coincidental that the poorest therapeutic responses and the single therapeutic failure observed to date have occurred in patients with nausea and vomiting, and in whom aureomycin blood levels have tended to be low. Much work is needed to correlate dose, whether oral or parenteral, with blood level, and to determine what levels need to be obtained. The possibility that certain rickettsial strains are more resistant to, or may acquire resistance to, the drug also requires investigation.

The occurrence of relapses in 2 patients after therapy was stopped, together with the relatively slow response in others, suggests that, in the dosages used, the effect of the drug may have been to suppress rather than to eradicate the infecting agent. Unless it is found that larger doses of the drug have a much more prompt effect on the course of the disease, it would thus appear that recovery depends on the development by the patient of sufficient immunity to overcome the infection while the drug is exerting a suppressive action. In this series, there has been one instance, in a patient treated on the second day, in which the development of complement-fixing antibodies appeared to have been lessened by the use of the drug.

In the present study, evidence of drug toxicity has been sufficiently slight to justify the use of still larger amounts of aureomycin in the

treatment of future cases. Such a procedure may make it possible to obtain a more uniform and more prompt response

Summary

1. Results of aureomycin therapy of 19 Q fever patients are presented.
2. The results in a group of 4 patients treated by the intramuscular route with small doses were not considered satisfactory.
3. Of the remaining 15 patients, all treated orally, improvement occurred in 14 relatively promptly after commencement of therapy. The fifteenth patient, classified as a chronic case of Q fever, failed to respond even to large doses of the drug.
4. Relapses occurred in 2 patients in the orally-treated group following cessation of therapy. However, both patients became, and remained, afebrile following a second course of aureomycin.
5. It is concluded that oral aureomycin therapy offers considerable promise in the treatment of Q fever and that further investigations are desirable to evaluate the usefulness of the drug.

References

1. WONG, S. C., & H. R. COX. 1948. Action of aureomycin against experimental rickettsial and viral infections. *Ann. N. Y. Acad. Sci.* 51(2): 290-305.
2. SHEPARD, C. C., & R. J. HUEBNER. 1948. Q fever in Los Angeles County. Description of some of its epidemiological features. *Am. J. Pub. Health* 38: 781-788.
3. LENNETTE, E. H., & G. MEIKLEJOHN. 1948. Q fever in central and northern California. *California Med.* 69(3): 197-199.

OCEAN SURFACE WAVES*

Consulting Editor: BERNHARD HAURWITZ

CONTENTS

The Solitary Wave and Periodic Waves in Shallow Water. By JOSEPH B. KELLER.....	345
The Action of Floating Bodies on Ocean Waves. By FRITZ JOHN....	351
The Breaking of Waves in Shallow Water. By JAMES J. STOKER, JR. 360	
The Solitary Wave Theory and Its Application to Surf Problems. By WALTER H. MUNK.....	376
Wave Motion at the Surface of a Current Which Has an Exponential Distribution of Vorticity. By ABDUL JABBAR ABDULLAH.....	425
Determination of the Depth of an Underwater Explosion from Measurements of the Dome of Spray. By CHAIM L. PEKERIS.....	442
Notes on Surface Waves. By MAURICE EWING AND FRANK PRESS..	453
The Propagation of Small Surface Disturbances Through Rotational Flow. By PHILIP DUNCAN THOMPSON.....	463
Recent Studies of Waves and Swell. By G. E. R. DEACON.....	475
Sea Surface Roughness Measurements in Theory and Practice. By H. R. SEIWELL.....	483
A Comparison Between Recorded and Forecast Waves on the Pacific Coast. By J. D. ISAACS AND THORNDIKE SAVILLE, JR.....	502
Variability in Direction of Wave Travel. By R. S. ARTHUR.....	511
Ocean Wave Research and Its Engineering Applications. By M. A. MASON.....	523
Details of Shore-based Wave Recorder and Ocean Wave Analyzer. By A. A. KLEBBA.....	533
Utilization of Wave Forecasting in the Invasions of Normandy, Burma, and Japan. By C. C. BATES.....	545

* This series of papers is the result of a Conference on Ocean Surface Waves held by the Section of Oceanography and Meteorology of The New York Academy of Sciences on March 18 and 19, 1948.
Publication made possible by a grant-in-aid from the Woods Hole Oceanographic Institution, Woods Hole, Massachusetts and by the General Funds of The New York Academy of Sciences.

Copyright, 1949. by The New York Academy of Sciences

THE SOLITARY WAVE AND PERIODIC WAVES IN SHALLOW WATER*

By JOSEPH B. KELLER

Institute for Mathematics and Mechanics, New York University, New York, N. Y.

In 1844, Scott-Russell¹ reported his experimental observations on the solitary wave, a wave consisting of a single elevation which propagates without change of form (the wave motion is in two dimensions only). Later, Boussinesq² and Rayleigh³ independently gave derivations of the approximate form and velocity-amplitude relationship for such a wave in shallow water. Then, by a slight modification of Rayleigh's method, Korteweg and DeVries⁴ obtained periodic waves of permanent type, which include the solitary wave as a special case when the wave length becomes infinite. Gwyther⁵ obtained Rayleigh's results by a slightly different method, and McCowan⁶ found more accurate results by "guided guessing." A systematic procedure for determining the wave velocity has been developed by Weinstein.⁷ The existence of the solitary wave has not yet been proved mathematically.

All the results mentioned above apply to irrotational, two-dimensional motion of an incompressible, inviscid fluid over a horizontal bottom. Rayleigh's treatment involves an iteration scheme of a peculiar kind and leads to a differential equation for the wave profile. The theory given by Boussinesq involves a number of physical assumptions in addition to those of the basic hydrodynamical theory. It also leads to a differential equation for the wave profile. Both of these methods assume that the depth of the water is small compared to some horizontal dimension, and they might be interpreted as developments of the whole problem in powers of the ratio of the depth to some horizontal dimension, such as wave length. However, because these procedures are so unsystematic, it is not clear that they are equivalent to such developments, nor to what order of approximation the solutions obtained are valid. The method of proceeding to higher approximations is also obscure.

The object of the present investigation is to discuss waves of permanent type† in shallow water by a method in which the character of the approximation is quite clear, and which is capable of being carried out to include terms of any desired order. The method consists in expanding the solution of the exact hydrodynamic problem systematically in powers of a dimensionless parameter $\sigma = (\omega h)^2$ where h is the depth of the undisturbed fluid and ω is the curvature at some point on the surface. The expansions are inserted

*The mathematical methods and results on which this paper is based are contained in "The Solitary Wave and Periodic Waves in Shallow Water" by Joseph B. Keller, Communications on Applied Mathematics, Inst. for Math. and Mech., NYU, Vol. I, No. 3, July, 1948.

†The author wishes to express appreciation to Professors K. O. Friedrichs and J. J. Stoker for their helpful discussions of this and related subjects.

‡Actually stationary solutions are found. Progressive waves may be obtained from them by adding a constant velocity to the fluid.

into the equations of motion and the boundary conditions, and coefficients of like powers of σ are equated. The variables are chosen in such a way that the terms of zero order in the expansion in powers of σ satisfy the well-known equations of the nonlinear shallow water theory, which are analogous to the equations of gas dynamics.*

It is easily shown that the only solutions of these equations for the first approximation (satisfied by the terms independent of σ) which are of permanent form are the constant or piecewise constant (shock type) solutions. However, the equations for the second approximation (satisfied by the coefficients of σ) have solutions which yield periodic waves of permanent form, similar to those of Korteweg and DeVries, and also solitary waves, similar to those of Rayleigh and Boussinesq.

The solution of permanent form, as given by the first and second approximations, for the irrotational, two-dimensional motion of an incompressible, inviscid fluid of mean depth h over a horizontal bottom is

$$\begin{aligned}\eta &= \eta_{\min} + (\eta_{\max} - \eta_{\min})cn^2 \left[\frac{x}{\lambda} 2F_1(k) \right], \\ p &= \rho g(\eta - y), \\ \bar{v} &= 0, \\ \frac{\bar{u}}{\sqrt{gh}} &= \frac{\eta_{\max}}{h} - L - \left(\frac{\eta_{\max}}{h} - \frac{\eta_{\min}}{h} \right) cn^2 \left[\frac{x}{\lambda} 2F_1(k) \right], \\ \frac{\lambda}{h} &= \frac{4}{\sqrt{3}} F_1(k) \left(2L + 1 - \frac{\eta_{\min}}{h} \right)^{-1/2}.\end{aligned}\tag{1}$$

In these equations, η is the surface elevation, measured up from the bottom; x is the horizontal coordinate and y is the vertical distance above the bottom; η_{\max} and η_{\min} are the maximum and minimum surface elevations, respectively; λ is the wave length, p the pressure, ρ the fluid density, g the acceleration of gravity, \bar{v} the vertical velocity, \bar{u} the horizontal velocity; F_1 and E_1 the complete elliptic integrals of the first and second kinds, respectively, of modulus k ; cn the Jacobi elliptic function of modulus k ; L and h are certain parameters given by

$$\begin{aligned}0 < k^2 &= \frac{\frac{\eta_{\max}}{h} - \frac{\eta_{\min}}{h}}{2L + 1 - \frac{\eta_{\min}}{h}} \leq 1, \\ \left(2L + 1 - \frac{\eta_{\min}}{h} \right) E_1(k) &= \left(2L - 2 - \frac{\eta_{\max}}{h} - \frac{\eta_{\min}}{h} \right) F_1(k).\end{aligned}$$

This solution yields a two-parameter family of periodic waves as does that

* These equations were first derived in this way by K. O. Friedrichs,* to whom this method is due.

of Korteweg and DeVries. The parameters η_{\max} and η_{\min} are subject only to the conditions

$$0 \leq \frac{\eta_{\min}}{h} \leq 1, \quad \frac{\eta_{\max}}{h} \geq 2 - \frac{\eta_{\min}}{h}, \quad (2)$$

which both follow from the definition of h or the relations among η_{\max} , η_{\min} , L , and k . For any values of η_{\max} and η_{\min} satisfying the inequalities above, the surface profile is periodic in x with the wave length λ . The height of the crest above the mean height is greater than the depth of the trough below the mean height. The crest is also narrower than the trough. The wave is, thus, not symmetric about the mean height as it is in the linear theory. This is shown in FIGURE 1, where one wave length of a typical wave profile is plotted.

FIGURE 2 shows the contour lines of $\frac{\lambda}{h}$ as a function of $\frac{\eta_{\min}}{h}$ and $\frac{\eta_{\max}}{h}$.

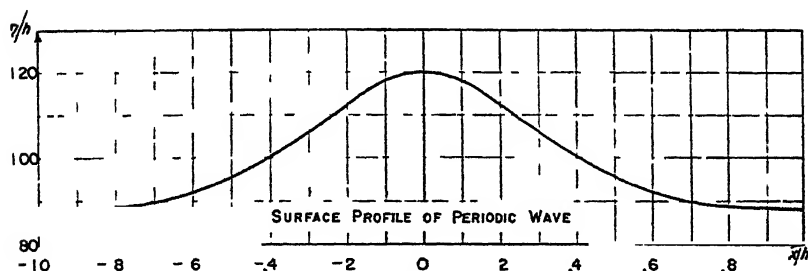


FIGURE 1.

From this figure, it can be seen that λ increases as either η_{\min} or η_{\max} increases. The qualitative behavior of this function is the same as that of the function obtained by Korteweg and DeVries, but quantitatively it is slightly different. Thus, although formally the equation for the wave profile obtained by Korteweg and DeVries is the same as the first of EQUATIONS 1, the two profiles are slightly different because of the different expressions for $\lambda(\eta_{\max}, \eta_{\min})$. Another difference between their solution and EQUATIONS 1 occurs in the expression for u , which they find to be independent of both x and y . From considerations of conservation of mass, that is much less reasonable than the dependence on x which is given by the fourth of EQUATIONS 1. The second of EQUATIONS 1 yields the pressure in the fluid, which is seen to be given by the hydrostatic expression even in the second approximation. This expression therefore seems to be very accurate. In the Korteweg-DeVries solution the pressure is not obtained.

From FIGURE 2, it can be seen that, for a fixed value of $\frac{\eta_{\min}}{h}$, as $\frac{\eta_{\max}}{h}$ decreases, λ decreases until $\lambda = 0$ when $\frac{\eta_{\max}}{h} = 2 - \frac{\eta_{\min}}{h}$. In this limiting case,

the amplitude may be finite, but the wave length is zero, so that the surface is everywhere discontinuous. On the other hand, if $\frac{\eta_{\max}}{h}$ increases indefi-

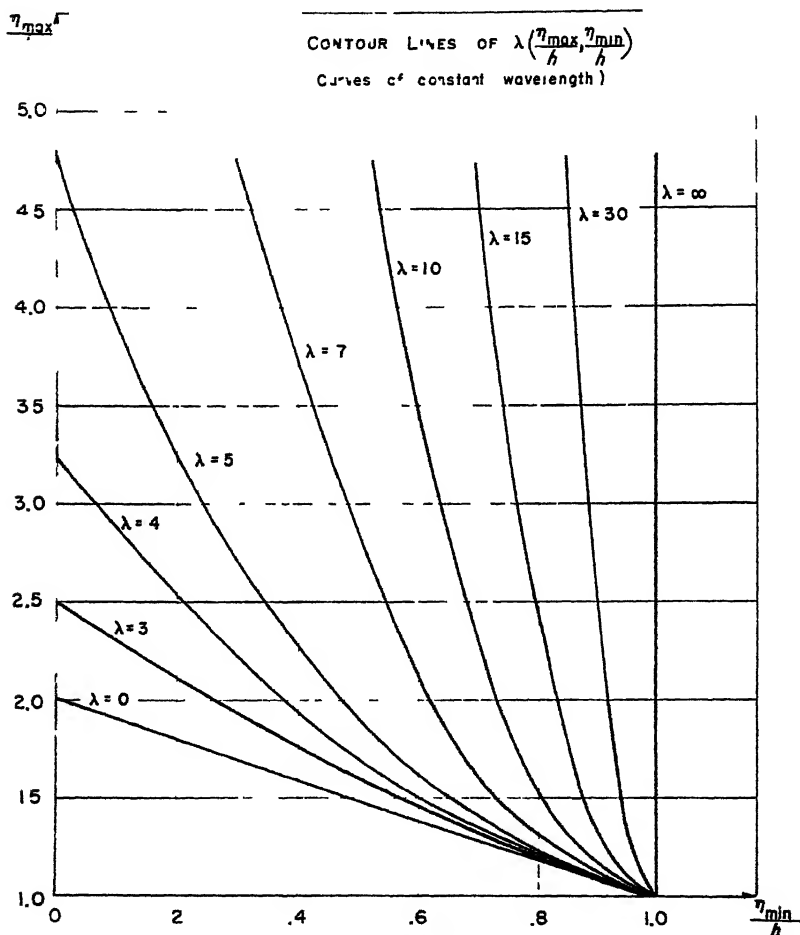


FIGURE 2.

nately, so does λ . While this occurs, the troughs become wider and the crests higher and relatively narrow.

It was hoped that the solution would impose an upper limit on $\frac{\eta_{\max}}{h}$, and that for the limiting solution the surface slope would be discontinuous at the crests. Then, as shown by Stokes,* an angle of 120° would be formed at the

* P. 418^o.

crests. However, to the present order of approximation these results are not obtained. Further research leading to these results would certainly be worth while.

When $\frac{\eta_{\max}}{h}$ and $\frac{\eta_{\min}}{h}$ are nearly equal to one, both the Korteweg-DeVries solution and EQUATIONS 1 reduce to the cosine solution of the linear shallow water theory. When $\frac{\eta_{\min}}{h} = 1$ and $\frac{\eta_{\max}}{h}$ is greater than one, but otherwise unrestricted, the wave length becomes infinite in both solutions. In this

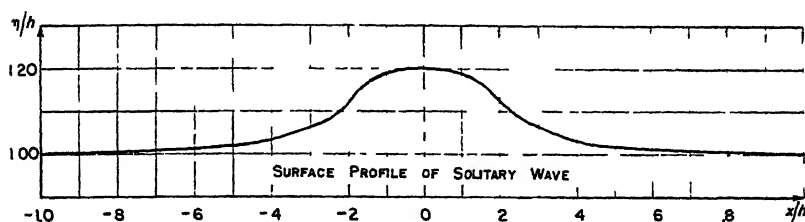


FIGURE 3

case, EQUATIONS 1 become (see FIGURE 3 for wave profile)

$$\begin{aligned}\eta &= h + (\eta_{\max} - h) \operatorname{sech}^2 \frac{x\sqrt{3}}{2h} \left(\frac{\eta_{\max}}{h} - 1 \right)^{1/2}, \\ p &= \rho g(\eta - \gamma), \\ \bar{v} &= 0, \\ \frac{\bar{u}}{\sqrt{gh}} &= \frac{1}{2} \left(\frac{\eta_{\max}}{h} + 1 \right) - \left(\frac{\eta_{\max}}{h} - 1 \right) \operatorname{sech}^2 \frac{x\sqrt{3}}{2h} \left(\frac{\eta_{\max}}{h} - 1 \right)^{1/2} \\ \lambda &= \infty.\end{aligned}\tag{3}$$

EQUATIONS 3 represent a solitary wave. The profile, given by the first of EQUATIONS 3, is the same as that found by Boussinesq.² The Korteweg-DeVries solution reduces, in this case, to the solitary wave found by Rayleigh. This solution, for the profile, differs from the first of EQUATIONS 3 only by having the additional factor $\sqrt{\frac{h}{\eta_{\max}}}$ multiplying the argument of the hyperbolic secant. For waves of small amplitude this factor is nearly one, and, thus, in this case, the Korteweg-DeVries-Rayleigh solution for the profile practically agrees with the first of EQUATIONS 3. Both the Korteweg-DeVries-Rayleigh solution and the Boussinesq solution yield a horizontal velocity independent of x and y . As mentioned above, considerations of conservation of mass indicate that the dependence of the velocity on x as given by the fourth of EQUATIONS 3 is more reasonable than the constant velocity. However, for x infinite, the fourth of EQUATIONS 3 yields $\bar{u} =$

$\sqrt{gh} \frac{\frac{\eta_{\max}}{h} + 1}{2}$, and the velocity $\sqrt{g\eta_{\max}}$, given by Korteweg-Devries-

Rayleigh, agrees with this to first order in the relative amplitude $\frac{\eta_{\max} - h}{h}$.

Thus, if the water at infinity is at rest, the propagation speeds of the wave, as given by the two solutions, agree to first order in the relative amplitude of the wave. However, according to the fourth of EQUATIONS 3, the water under the crest would be moving, while, in the Korteweg-DeVries-Rayleigh solution, it would be stationary.

Bibliography

1. SCOTT-RUSSELL. 1844. Report on Waves. British Association Reports.
2. BOUSSINESQ. 1871. (C. R. June 19.)
3. RAYLEIGH. 1883. The form of standing waves on running water. Proc. London Math. Soc. XV: 69.
4. KORTEWEG & DEVRIES. 1895. On the change of form of long waves advancing in a rectangular canal and on a new type of long stationary waves. Phil. Mag., XXXIX (5): 422.
5. GWYTHER. 1900. Phil. Mag. L (5): 213, 308, 349.
6. McCOWAN. 1891. On the solitary wave. Phil. Mag. XXXVII (5): 45; 1894. On the highest wave of permanent type, Phil. Mag. XXXVIII (5): 351.
7. WEINSTEIN, A. 1926. Lincei III (6): 463.
8. LAMB, H. 1945. Hydrodynamics: 423-427. Dover.
9. FRIEDRICHS, K. O. 1948. On the derivation of the shallow water theory, appendix to the formation of breakers and bores by J. J. Stoker. Communications on Applied Mathematics. Inst. of Math. and Mech. N. Y U I: (1).

THE ACTION OF FLOATING BODIES ON OCEAN WAVES

By FRITZ JOHN

Institute for Mathematics and Mechanics, New York University, New York, N. Y.

In this paper, I want to present some results concerning the motion of a liquid in the presence of an obstacle floating on the surface.* The points I want to discuss are the exact mathematical formulation of the problem, the stability of equilibrium of floating bodies, the behavior of waves at infinity, and, finally, explicit solutions obtained for some special situations.

The state of a liquid, under the assumptions of irrotationality and incompressibility, can be described by means of two functions, the velocity potential $\phi(x, y, z, t)$ and the pressure p . Here ϕ satisfies Laplace's equation, and p is expressible in terms of ϕ by means of Bernoulli's equation.

Along any boundary surface of the liquid, we have the kinematic condition, expressing that the normal derivative of ϕ is equal to the normal velocity of the boundary surface, and the dynamic condition, that the pressure is continuous across the boundary surface. This dynamic condition is applied along the "free surface" of the liquid as the condition that the pressure is constant and equal to the atmospheric pressure. Along the immersed surface of a floating body, the dynamic condition is used by expressing that the only forces acting on the body are gravity and the pressure of the liquid on the immersed surface.

There is little one can do about determining the motion in this generality. The principal mathematical difficulty lies in the fact that we have boundary conditions for ϕ along boundaries that vary with the time. The first step usually taken to simplify the conditions is to "linearize" the problem by restricting oneself to infinitesimal motions of liquid and body about average positions. One then obtains conditions for the potential function ϕ along the "average" boundary surfaces, which are independent of the time, and correspond to the "rest" or "equilibrium" position of the system.

Using a coordinate system, in which the mean free surface coincides with the xy -plane, the boundary condition for ϕ on the mean free surface takes the classical form

$$\phi_{tt} + g\phi_n = 0.^\dagger$$

The position of the rigid floating body may be described by the coordinates X, Y, Z of its center of gravity and by the components $\theta', \theta'', \theta'''$ of the infinitesimal rotation about the center of gravity that carries the body B from its rest position into the actual position at the time t . In the rest position, the center of gravity may be at X^0, Y^0, Z^0 , and $\theta', \theta'', \theta'''$ vanish.

* These results were obtained in connection with investigations pursued under a contract with the Office of Naval Research of the U. S.

† Subscript is used to denote partial derivatives.

If S denotes the immersed surface of B at the time t , and S^0 the immersed surface in the rest position, we have the linearized kinematic condition

$$\begin{aligned}\frac{\partial \phi}{\partial n} = & X_t \frac{\partial x}{\partial n} + Y_t \frac{\partial y}{\partial n} + Z_t \frac{\partial z}{\partial n} + \theta'_t \left[(y - Y^0) \frac{\partial z}{\partial n} - (z - Z^0) \frac{\partial y}{\partial n} \right] \\ & + \theta''_t \left[(z - Z^0) \frac{\partial x}{\partial n} - (x - X^0) \frac{\partial z}{\partial n} \right] \\ & + \theta'''_t \left[(x - X^0) \frac{\partial y}{\partial n} - (y - Y^0) \frac{\partial x}{\partial n} \right]\end{aligned}$$

valid along S^0 . Here $\frac{\partial}{\partial n}$ denotes differentiation in the direction of the normal of S^0 .

It is more difficult to write down the linearized version of the six dynamical equations which express that the rates of change of momentum and of

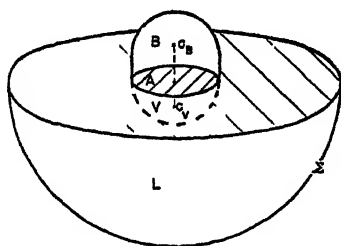


FIGURE 1

angular momentum of the body B are due to the pressure of the liquid on the immersed surface and to gravity. There are, first, the static conditions characterizing the equilibrium position: (1) Mass M of B = mass of displaced liquid V^0 (Archimedes law), (2) The center of gravity (X^0, Y^0, Z^0) of B and the center of gravity of the displaced liquid V^0 in the rest position lie on the same vertical line. (See FIGURE 1.) The six dynamical equations involving the time derivatives of the quantities $X, Y, Z, \theta', \theta'', \theta'''$ are rather complicated, and only one of them may be given here as an example, the one regulating the vertical motion of the body:

$$\begin{aligned}\frac{1}{\rho} M Y_{tt} = & - \iint_{S^0} \phi_t \frac{\partial y}{\partial n} dS - g(Y - Y^0) \iint_A dx dz \\ & - g\theta''' \iint_A (x - X^0) dx dz + g\theta' \iint_A (z - Z^0) dx dz.\end{aligned}$$

Here ρ is the constant density of the liquid, and A is the cross section of the

free surface $y = 0$ with the body in the equilibrium position. This formula puts into evidence that the pressure acting on B consists of a non-hydrostatic part contributed by ϕ_t and a hydrostatic part which is due to the deviation of the body B from its equilibrium position.

To complete the conditions for ϕ we still have to prescribe the behaviour at infinity and initial conditions. The conditions to be imposed at infinity are suggested by energy considerations. Taking an open surface Σ containing B and a portion L of the liquid, it is seen that the outward flow of energy through Σ in unit time is given by

$$-\rho \iint_{\Sigma} \phi_t \frac{\partial \phi}{\partial n} dS$$

when n is the exterior normal of Σ . Transforming this expression with the help of the boundary conditions mentioned, we find that it can be written in the form

$$\begin{aligned} -\frac{d}{dt} \left[\iiint_L \frac{\rho}{2} (\phi_x^2 + \phi_y^2 + \phi_z^2) dx dy dz + \frac{\rho}{2g} \iint_{y=0} \phi_t^2 dx dz \right. \\ \left. + \iint_B (\dot{x}^2 + \dot{y}^2 + \dot{z}^2) dm + \rho g Q(Y - Y^0, \theta', \theta''') \right]. \end{aligned}$$

Here, the second integral is extended over that part of the free surface $y = 0$ outside B . We can interpret this formula as representing the energy of the system ($B + L$) as the sum of 4 parts: (1) the kinetic energy of L ; (2) the potential energy of L due to the waves; (3) the kinetic energy of B ; (4) the potential energy of the system ($B + L$) due to the deviation of the immersed surface from its equilibrium position. The first 3 parts are, by their nature, positive definite quadratic forms. The last part ($\rho g Q$) is a quadratic form in those variables that affect the buoyancy of B .

This identity for the flow of energy through the surface Σ immediately furnishes a stability criterion for the equilibrium position. If Q is a positive definite form, then the whole energy expression is non-negative at all times, and is, besides, a decreasing function of the time, if no energy flows from the outside to the inside of Σ . Thus, in the absence of exterior sources of energy, the total energy of the system of liquid and floating body inside Σ will never exceed its initial value, hence the maximum possible values of $|Y - Y^0|$, $|\theta'|$, $|\theta'''|$ are bounded by the initial disturbance. Thus, the body cannot be upset by a small disturbance. In this way, one recognizes that the positive definite character of Q is a sufficient condition for stability of the equilibrium position. Interpreting this condition geometrically we are led to the classical stability condition:

The equilibrium position is stable, if the product of the volume of the displaced liquid V and of the height of the center of gravity of B above that

of V does not exceed the moment of inertia of the area A about any horizontal axis.

This condition has usually been derived from static considerations, which neglect the motion of the liquid.* Taking into account the motion of the liquid, we see that the condition is only sufficient for stability, as the argument given here requires only that the sum of all 4 terms in the energy expression be positive definite.

Another consequence of the identity for the flux of energy is the following uniqueness theorem:

Let no external sources of energy be present, and let the position of the floating body be close to a position of equilibrium. Then the motion of liquid and body are uniquely determined, if we prescribe (a) initial positions and velocities of all particles, (b) the wave motion at infinity, except for outgoing waves (*i.e.* waves carrying energy outwards to infinity).

Thus, at infinity, only the "non-outgoing" part of the waves can be prescribed. Mathematically, it is not obvious how a wave is to be split up into an outgoing and non-outgoing part. This decomposition can, however, be carried out explicitly in the important special case of motions in a liquid of infinite depth, which are periodic and simple harmonic functions of the time. In this case, the velocity potential ϕ satisfies an identity

$$\phi_{zz} = -k^2\phi,$$

and the free surface condition takes the simple form

$$\phi_y = \lambda\phi,$$

where $\lambda = k^2/g$. One can then prove rigorously that ϕ can be split up in a unique fashion into two parts:

$$\phi = \phi_1 + \phi_2.$$

Here ϕ_1 is an everywhere regular solution, *i.e.* it corresponds to a wave motion, as it would take place in the absence of the floating body. ϕ_2 is a combination of purely outgoing waves, the secondary waves set off by the obstacle. ϕ_1 can be written in the form

$$\phi_1 = \text{Real part} \left(e^{-ikz + \lambda y} \sum_{n=-\infty}^{+\infty} c_n e^{-in\theta} J_n(\lambda R) \right),$$

where the c_n are complex constants, and $x = R \cos \theta$, $z = R \sin \theta$. Thus, ϕ_1 is an exponential function in the depth $-y$. ϕ_2 permits a similar expansion at infinity, but not at finite distances:

$$\phi_2 = \text{Real part} \left(e^{-ikz + \lambda y} \sum_{n=-\infty}^{+\infty} d_n e^{-in\theta} H_n^{(1)}(\lambda R) \right) + 0 \left(\frac{1}{x^2 + y^2 + z^2} \right)$$

* See Appell's "Mécanique Rationnelle," 5: 213, 208

J_n and $H_n^{(1)}$ are respectively the Bessel function and Hankel functions of order n . The outward flow of energy through a far away surface Σ is given by

$$4k \sum_{n=-\infty}^{+\infty} (c_n \bar{d}_n + d_n \bar{c}_n + \bar{d}_n d_n).$$

It is certainly positive in the absence of the "everywhere regular" part ϕ_1 . The secondary waves (ϕ_2) die out at infinity, whereas this is not necessarily the case for the regular solution ϕ_1 . If we put $\phi_2 = \text{Re}(\psi)$, then ψ_2 satisfies the Sommerfeld condition $\lim_{R \rightarrow \infty} \left(\frac{\partial}{\partial R} \psi_2 - i\lambda \psi_2 \right) = 0$

One arrives in this way at a natural formulation of the complete data by which the wave motion is to be determined, at least in the case of periodic motions. One prescribes arbitrarily the primary part (ϕ_1) of the waves, say $\phi_1 = \text{Real part}(\psi_1)$. The determination of the secondary waves, $\phi_2 = \text{Real part}(\psi_2)$, can then be reduced to a standard mathematical problem, the solution of a Fredholm integral equation. For this purpose, one constructs a suitable Green's function Γ_P corresponding to a secondary wave issuing from an arbitrary point P of the liquid. One then obtains an integral equation

$$2\pi i \psi_1(P) = - \int \int_{S^0} \left[(\psi_1 + \psi_2) \frac{\partial \Gamma_P}{\partial n} - \Gamma_P \frac{\partial (\psi_1 + \psi_2)}{\partial n} \right] dS$$

connecting the values of ψ_2 and ψ_1 and their normal derivatives on S^0 . As the normal derivative of ψ_2 on S^0 is expressible in terms of the velocity components of the body B , one obtains, by solving this integral equation, ψ_2 as a linear combination of the primary wave ψ_1 and of $X_t, Y_t, \dots, \theta_t'''$. Substituting the resulting expression for ϕ into the dynamic equations for the body motion, we arrive at a system of linear algebraic equations for $X_{tt}, Y_{tt}, \dots, \theta_{tt}'''$, determining the body motion. Thus, the whole problem of determining the motion from given primary waves is reduced to the solution of a Fredholm integral equation.

In conclusion, I want to present certain concrete numerical results that have been obtained for the case of shallow water. The liquid is here supposed to have constant depth h . The waves considered shall have a wave length which is large compared with the depth h . At the same time, the floating body is supposed to be very "flat," in the sense that its radius of curvature shall be large compared with the depth h . I shall restrict myself to the case of two-dimensional motion, where the body B is taken as a floating cylinder with horizontal generators and with flat cross section.

By a formal expansion scheme, very similar to the one used by Dr. Keller in the preceeding talk, one then arrives at the wave equation

$$\phi_{tt} = g h \phi_{zz}$$

in points under the free surface. If $y = \eta(x)$ is the equation of the cross section of the cylinder in the rest position, one has, for values x corresponding to points under the cylinder, the differential equation

$$\frac{\partial(h + \eta)\phi_x}{\partial x} = -Y_t + \eta_x X_t - (x - X^0)\theta_t''$$

which essentially expresses the law of preservation of mass. The dynamical equations give here 3 simple integral relations. These relations can all be solved explicitly in terms of given initial conditions and given incoming waves. The body position here has, of course, only 3 degrees of freedom, represented by the variables X, Y, θ'' . The calculations have been carried out for certain typical cases, which I want to discuss here, by Mr. Isaacson and Miss Johnson of the Institute for Mathematics and Mechanics at New York University.

The first case considered may be that of a cylinder symmetrical to a vertical plane. This cylinder shall have been raised above its equilibrium position by a distance A and then released. The resulting motion depends on a certain "draught-coefficient" δ , which roughly measures the depth to which the cylinder B is immersed compared with the depth of the liquid under the cylinder. Thus, δ varies from the value 0 for cylinders which are only very slightly immersed, to the value ∞ for cylinders resting on the bottom.*

It turns out that the resulting motions of body and liquid are given by damped harmonic functions of the time. The wave profile for $\delta = 2$, after two time periods, is represented on FIGURE 2 (the horizontal distances are foreshortened in that FIGURE, so that the body does not appear as "flat" as the theory demands), and the amplitude of the body motion is given in FIGURE 3. The damping is in general very strong, so that the waves generated by the body motion appear to consist of a single half-sine-wave propagating with speed \sqrt{gh} . The coefficient of damping during one half-period is given by the formula

$$\gamma = e^{-\pi \sqrt{1+4\delta}}$$

Thus, e.g., $\gamma = .0043$ for $\delta = 0$, $\gamma = .16$ for $\delta = 2$, $\gamma = 1$ for $\delta = \infty$. This means that a slightly immersed body raised an (infinitesimal) distance A above its equilibrium position, will only sink a maximum distance $.0043A$ below its equilibrium position in the resulting motion; after half a period, it will be practically at rest.

* The exact definition of δ is

$$\delta = \int_{g^0} \frac{h x^2}{h + \eta} dx / \int_{g^0} x^2 dx$$

The wave length of the resulting waves depends on the width D of the body at the water line. It is given by

$$L = \frac{2\pi(1 + \delta)}{\sqrt{3(1 + 4\delta)}} D.$$

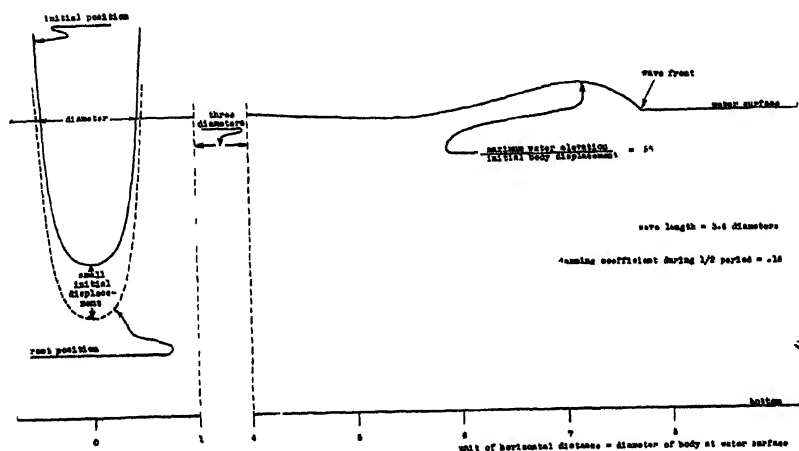


FIGURE 2. Wave profile after two full periods generated by initial vertical displacement of floating cylinder submerged (at rest) to about $\frac{2}{3}$ of depth of water.

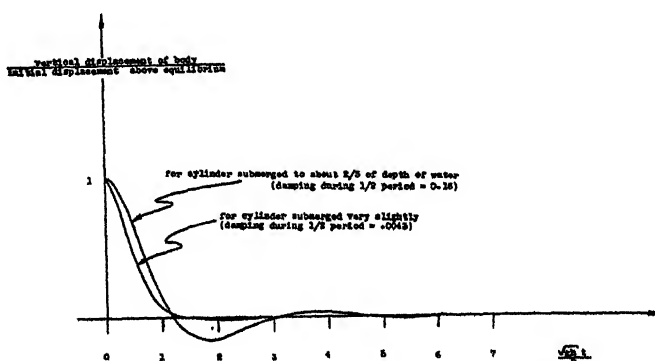


FIGURE 3. Body motion.

It varies from $L = 3.6D$ for $\delta = 0$, has its least value $L = 3.1D$ for $\delta = \frac{1}{2}$, and becomes arbitrarily large for $\delta \rightarrow \infty$. The height of the leading wave is given by the expression

$$A \sqrt{\frac{3}{1 + \delta}} e^{-\sqrt{\frac{3}{1 + 4\delta}} \text{arc tan } \sqrt{\frac{1 + 4\delta}{3}}},$$

and has, e.g., the value $.7A$ for $\delta = 0$.

The second case considered is that of the motion of a cylinder generated by a periodic wave of amplitude α and wave length L coming in from infinity. In this case, the cylinder, again assumed to be symmetrical, will be displaced laterally, vertically, and will also rotate. The amplitude of

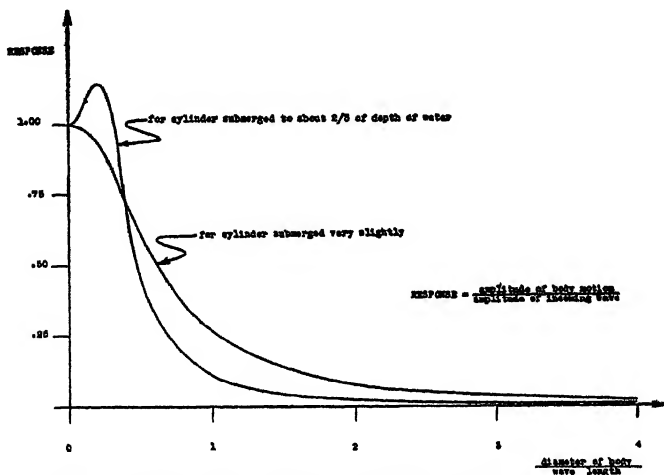


FIGURE 4. Periodic motion of cylinder generated by periodic incoming wave.

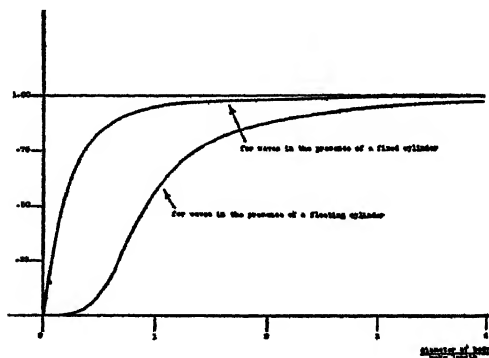


FIGURE 5. Reflection coefficients.

of the vertical displacement depends on the ratio of L and D , and also on the draught coefficient δ :

$$A = \frac{\alpha}{\sqrt{\left[1 - \frac{\pi^2 D^2}{3L^2} (1 + \delta)\right]^2 + \frac{\pi^2 D^2}{L^2}}}$$

The formula shows that for small D and δ the coefficient is approximately 1; thus, a small body simply follows the motion of the liquid in its vertical

motion.* On the other hand, the coefficient is 0 for large D or δ , *i.e.* a very large body is only slightly excited by the incoming waves. There appears to be a certain resonance phenomenon for bodies with draught coefficient exceeding $\frac{1}{2}$. For such bodies, there are wavelengths for which the amplitude of the vertical body motion exceeds that of the incoming wave. (See FIGURE 4.)

The incoming wave is partly reflected and partly transmitted. The reflection coefficient has been calculated for the case of a cylinder, which is only slightly immersed, and has approximately rectangular cross section. FIGURE 5 represents the graph of the reflection coefficient for various values of D/L . For comparison, there has been drawn the reflection coefficient for a wave reflected on an obstacle of the same shape, but held fixed, instead of floating freely. In both cases, there is practically no reflection on very small objects (D/L near 0), and almost complete reflection on a very large object (D/L large). In every case, the reflection is less, if the object is floating, and hence yielding to the wave. For small objects, the reflection coefficient is even of a different order of magnitude than for fixed obstacles. Reflection on a fixed obstacle is then proportional to the width D of the obstacle, whereas, for a floating one, it is proportional to the fifth power of the width.

The preceding numerical results have not yet been compared with experiments. It should be simple to test them experimentally, as most of the quantities involved should be easily observable.

* The same can be seen for the lateral motion and for the rotation.

THE BREAKING OF WAVES IN SHALLOW WATER

By JAMES J. STOKER, JR.

Institute for Mathematics and Mechanics, New York University, New York, N. Y.

Introduction

The purpose of this paper is to discuss the breaking of waves along a shallow beach from the mathematical point of view. The theory used is an approximate theory derived from the basic hydrodynamics of irrotational non-viscous flow. It is generally referred to as the shallow water theory. The theory seems to have been derived long ago (Lagrange, 1781). This theory is accurate only for waves with wave lengths large compared with the depth of the water, but it is not necessary for its accuracy to assume that the wave amplitudes are small. The long wave-shallow water theory is mathematically analogous to the theory of compressible flow of a gas, and, as a consequence of the interest in aerodynamics, this theory has been developed very extensively. One of the principal objects of this paper is to interpret certain well known results in gas dynamics in terms of water wave phenomena. In particular, the development of a discontinuous shock wave in a compressible gas has as its analogue the development of a breaker in water.

After a brief statement of the mathematical theory in the section on non-linear shallow water theory, there is formulated the method of characteristics which is used to solve the problems. In the section on development of breakers and bores, the method of characteristics is used to explain why waves break in shallow water. Potentially, the most useful application of the theory presented here is probably to the problems encountered in studying the progress of flood waves or surges of other kinds down a river. For a discussion of these problems, as well as a detailed development of the general mathematical theory, reference is made to a previously published paper of the author (Stoker, 1948).

Nonlinear shallow water theory

In the book of Lamb (1945) the shallow water theory is derived from the assumption that the pressure p in the water is given by the hydrostatic law

$$p = g\rho(\eta - y), \quad (1)$$

in which η is the height of the water above the undisturbed level and ρ is the density of the water (see FIGURE 1). A consequence of this assumption is that the x -component u of the velocity of the water particles does not depend on the depth coordinate y , hence $u = u(x, t)$, since we assume always that the motion is parallel to the x, y -plane. The surface elevation

η and the velocity u are then readily found to satisfy the differential equations

$$\eta_t + [u(\eta + h)]_x = 0,^* \quad (2)$$

$$u_t + uu_x + g\eta_x = 0. \quad (3)$$

The quantity h represents the undisturbed depth of the water.

If we introduce the following new quantities:

$$\bar{p} = \int_{-\eta}^{\eta} p \, dy, \quad \bar{\rho} = \rho(\eta + h), \quad (4)$$

we find from (1) that \bar{p} and $\bar{\rho}$ are related as follows:

$$\bar{p} = \frac{g\rho}{2} (\eta + h)^2 = \frac{g}{2\rho} \bar{\rho}^2; \quad (5)$$

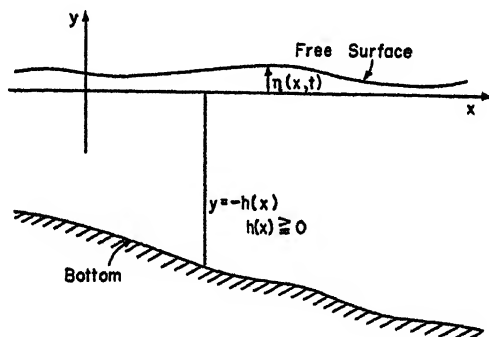


FIGURE 1.

and that the differential EQUATIONS 2 and 3 take the form

$$(\bar{\rho}u)_x = -\bar{\rho}_t, \quad (6)$$

$$\bar{\rho}(u_t + uu_x) = -\bar{p}_x + g\bar{\rho}h_x. \quad (7)$$

In the special case in which the undisturbed depth of the water is constant, so that $h_x = 0$, EQUATIONS 6 and 7 are the differential equations for unsteady one-dimensional motion of a compressible gas of pressure \bar{p} and density $\bar{\rho}$ which satisfies the "adiabatic law" (5) with the fixed adiabatic exponent 2. This analogy seems to have been pointed out first by Riabouchinsky (Riabouchinsky, 1932). It has been used (von Kármán, 1938, and Preiswerk, 1938) to study two-dimensional steady flows of water at relatively high speeds in open channels. The analogy is also used in the aircraft industry in connection with supersonic flow around obstacles (*cf.* Bruman, 1947, and Knapp and Einstein, 1946).

* Differentiations are indicated throughout this paper by letter subscripts.

The method of characteristics

We formulate the theory of characteristics for the case in which the depth h of the water is constant. In this theory, the quantity c , defined by

$$c = \sqrt{g(\eta + h)}, \quad (8)$$

is introduced in place of the surface elevation η . This quantity is the analogue of the sound speed in a gas. In connection with water waves, it determines the local velocity with which a small disturbance propagates. The characteristic theory is then formulated in the following equations:

$$\begin{cases} C_+: \frac{dx}{dt} = u + c, & u + 2c = k_1 = \text{const.} \\ C_-: \frac{dx}{dt} = u - c, & u - 2c = k_2 = \text{const.} \end{cases} \quad (9)$$

These relations are to be interpreted in the following way: Functions $u(x, t)$ and $c(x, t)$, which satisfy the original differential equations of our problem, lead to first order ordinary differential equations $\frac{dx}{dt} = u \pm c$ so that $u \pm 2c$ is constant along the respective solution curves of these differential equations. The two families of curves are called *characteristics*. More important for our purposes is the fact that the converse is also true, *i.e.*, if two families of curves in an x, t -plane satisfying $\frac{dx}{dt} = u \pm c$ and also $u \pm 2c = \text{const.}$ along these curves can be found, functions $u(x, t)$ and $c(x, t)$ are determined, and these functions satisfy the original partial differential equations.

There is an important class of problems in which one of the two families of characteristics is a set of straight lines. In fact, it can be shown that the following statement holds: *If there exists a straight line characteristic along which u and c are constant, then that characteristic is embedded in a whole family of straight characteristics along each of which u and c are constant.* Since the undisturbed state of rest of the water corresponds to $u = 0$, $c = \text{const.}$, it follows immediately from EQUATION 9 that the characteristics for such a state are straight lines; hence any continuous motion which develops from the state of rest is describable in terms of characteristics, one family of which is a set of straight lines. The study of such solutions, which have been called simple waves (Courant and Friedrichs, 1944), goes back to Earnshaw and Riemann in the middle of the last century.

We now consider the motion which results in water at rest when a disturbance is created at one point and allowed to propagate into still water. It is convenient to think of the water as filling a long tank, one end of which can be set in motion in order to create the disturbance. The problem which arises is the exact analogue of the problem of determining the motion of a gas

when a piston is pushed into a long tube filled with a compressible gas at rest. In FIGURE 2, we indicate the straight characteristics for a problem in which the end of the tank is given a velocity away from the water in the tank. At the time $t = 0$, the water is at rest and has the constant depth h_0 . The end A of the tank is accelerated to the left from rest so that it acquires a velocity u_A . The straight characteristics are shown in the x, t -plane,

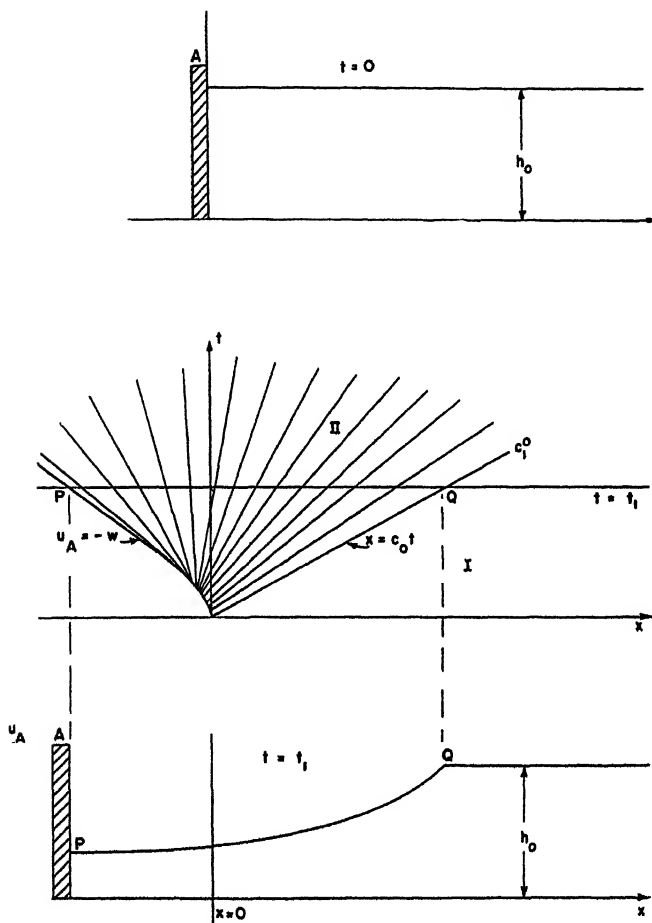


FIGURE 2.

which is divided into two zones, marked I and II, separated by the line $x = c_0 t$ with $c_0 = \sqrt{gh_0}$. The zone I is the zone of rest (the characteristics are not drawn in it). It is terminated by the line $x = c_0 t$, which means simply that, at a given point* x in the tank, no disturbance is felt until the

* It should be recalled, at this point, that the velocity of the water is, in the present theory, the same for all particles on the same vertical at any given instant. The state of the water depends only on the coordinate x and the time t .

time t given by $x = c_0 t$ elapses, which means, in turn, that no effect of the disturbance is noted until the wave initiated at $x = 0$ at the time $t = 0$ has time to reach the point. The zone II is the zone of disturbance. We indicate only the family of straight characteristics. The initial disturbance travels with the velocity c_0 of a small disturbance in water of the depth h_0 , as one might expect. At a later time, indicated by the line $t = t_1$ on the diagram of the straight characteristics, the corresponding shape of the water surface is indicated: the part corresponding to values of x between P and Q is the disturbed part in which the water surface is progressively lowered going to the left, while the surface is undisturbed to the right of Q . The method of determining the disturbance through use of the characteristics follows.

It is a relatively easy matter to calculate the slopes of the straight characteristics in the zone II emanating from the curve $x_A = x_A(t)$ determining the position of the left end of the tank, as well as the values of the velocity u and wave speed c , from which the depth is immediately determined and which belong to each of them (we recall that these quantities are constant along the straight characteristics), once the initial depth h_0 in the zone of rest and the velocity u_A of the water at the left end of the tank are given. For details, see the previously cited paper of the author (Stoker, 1948). At the time $t = t_1$, for example, the values of u and c for any given x are determined through the values of u and c assigned to the straight characteristic which crosses the line $t = t_1$ at the point with abscissa x .

Development of breakers and bores

The analogy of the shallow water theory with gas dynamics is perhaps most strikingly revealed through consideration of the problem of breaking of waves. It is of interest to recall how waves break on a shallow beach. FIGURES 3, 4, and 5 are photographs* of waves on the California coast. FIGURE 3 is a photograph from the air, taken by the Bureau of Aeronautics of the U. S. Navy, which shows how the waves coming from deep water are modified as they move toward shore. The waves are so smooth some distance off shore that they can be seen only vaguely in the photograph, but as they move in shore the front of the wave steepens noticeably until, finally, breaking occurs. FIGURES 4 and 5 are pictures of the same wave, with the picture of FIGURE 5 taken at a slightly later time than the previous picture. The steepening and curling over of the wave are very strikingly shown.

In FIGURE 2, we have indicated the character of a wave of depression which propagates into still water. The essential point, for our purposes now, is that such a wave propagates indefinitely as a smooth continuous wave as long as the movable left end of the tank is accelerated to the left, away from the water. If, however, the left end of the tank is pushed into the water to the right, in order to create a hump or elevation above the undis-

* These photographs were very kindly given to me by Walter Munk of the Scripps Institution of Oceanography.

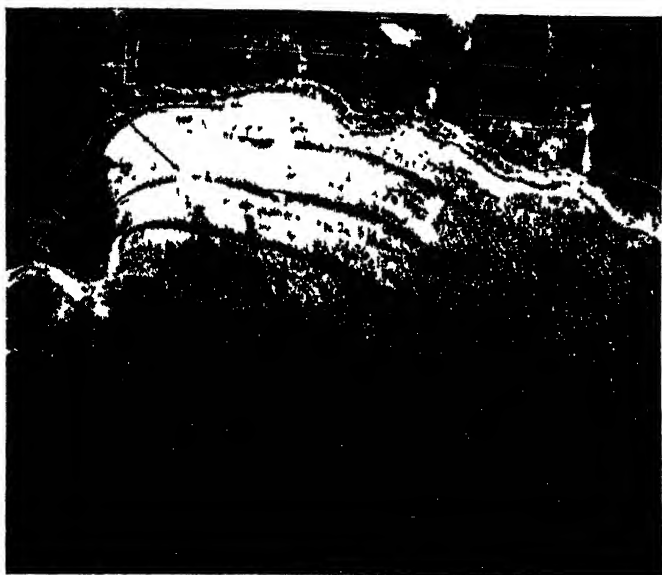


FIGURE 3



FIGURE 4

turbed level, the effect is very different from the mathematical point of view. In this case, the wave can advance as a smooth continuous wave only for a

limited distance and time, since no continuous solution of the differential equations exists beyond certain values of x and t . What happens in such a case is indicated in FIGURE 6, which shows the straight characteristics in an x, t -plane for a case of the type under discussion. The curve AD gives the position x of the left end of the tank as a function of t . In contrast with the situation shown in FIGURE 2, in which the end of the tank was moved to the left, we observe that the straight characteristics no longer diverge as they go out from the curve AD , but rather are turned toward each other so that they eventually intersect. This is a fact which can be readily deduced from the characteristic equations. In general, in cases of the type under

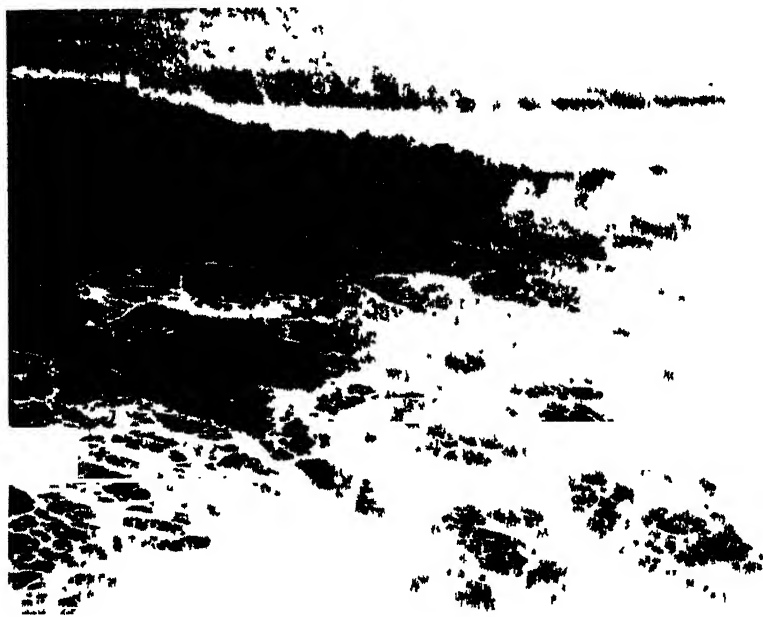


FIGURE 5

discussion, the characteristics will have an envelope indicated by the curves emanating from the point E in FIGURE 6. But since the values of u and c are different on different characteristics, it follows without much difficulty that continuous solutions of the differential equations will not exist for values of x and t , both larger than those at point E . In fact, the slope of the water surface, at the point (x_E, t_E) corresponding to point E , can be shown to be infinite. *We interpret this behavior of the solution of our mathematical problem to mean physically that the wave, having reached a state in which the front of the wave is vertical at one point, will shortly thereafter curl over and break.* In gas dynamics, the building up of a shock wave from a continuous compression wave is explained in the same way.

A simple qualitative explanation for the development of a breaker has often been given (cf., for example, an appendix by Harold Jeffreys to the book of Cornish, 1934). This explanation makes use of the fact (known from theory as well as observation) that the propagation speed of a wave increases with the height of the wave above the undisturbed water level. Consequently, if a wave is created in such a way as to cause a steady rise in the water surface, it follows that the higher points on the wave surface will propagate at higher speed than the lower points in front of them. In other words, there is a tendency for the higher portions of the wave to overtake and to crowd the lower portions in front, so that the front of the wave becomes steep and eventually curls over and breaks. On the other hand, a

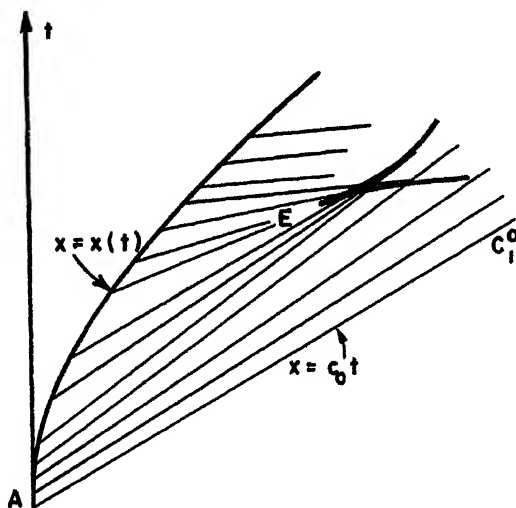


FIGURE 6

depression wave tends to be smoothed out, as one can see by a similar argument.

We give now a few results of numerical calculations for a few cases in which the "pulse" created at $x = 0$ and traveling into still water is a portion of a sine wave with elevation $\eta(0, t)$ given by

$$\eta(0, t) = A \sin \omega t. \quad (10)$$

FIGURE 7 indicates the straight characteristics for a case in which A is positive so that it is a hump which travels into still water. The point (x_b, t_b) marks the beginning of the envelope. We shall refer to this point as the breaking point, in accord with the remarks above. It is possible to calculate

the position of the breaking point as determined by our theory for a pulse given by (10). The result is

$$x_b = \frac{2c_0(c_0 + u_0)^2}{3gA\omega}, \quad (11)$$

$$t_b = \frac{2c_0(c_0 + u_0)}{3gA\omega}. \quad (12)$$

In these formulas, the quantity u_0 refers to a uniform initial velocity of the water assumed to exist at the instant the pulse given by (10) is initiated. If $u_0 = 0$, we note that the time and distance to the breaking point increase

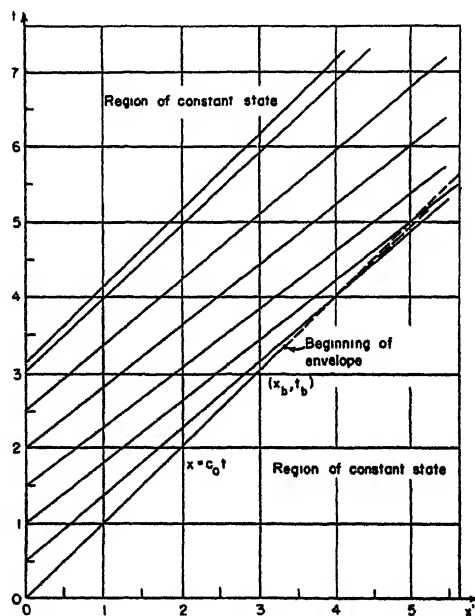


FIGURE 7

with the $3/2$ power and the first power of the initial depth respectively, when the amplitude A and "frequency" ω of the pulse are held fixed, since $c_0 = \sqrt{gh_0}$. Breaking occurs earlier when the amplitude and frequency are larger; hence short waves break sooner than long waves, since long waves are correlated with lower frequencies. The effect of an initial velocity u_0 of the whole body of water is also clear. Early breaking is favored by smaller values of u_0 . In fact, if u_0 is negative, *i.e.* if the water is flowing initially toward the point where the pulse originates, the breaking may occur very quickly. Everyone has observed this phenomenon at the beach, where the breaking of an incoming wave is often observed to be hastened by water rushing down the beach from the breaking of a preceding wave.

FIGURE 8 shows the shape of the water surface derived from the diagram of characteristics shown in FIGURE 7 at two different times. The full curve shows the pulse at the point which we call the breaking point. As we observe, the slope of the water surface is infinite at the front of the wave. FIGURE 9 was drawn using the characteristics as given by FIGURE 7 for a time considerably greater than the time of breaking t_b . The dotted part of the curve in FIGURE 9, showing the shape of the wave, was drawn using the region between the two branches of the envelope of the straight characteristics. This method is illegitimate mathematically, but may still have a certain significance physically, as indicating the process of curling over of the wave.

FIGURE 10 shows four stages in the shape of a wave (approximately a sine wave) moving into still water when the front portion of the wave is a depres-

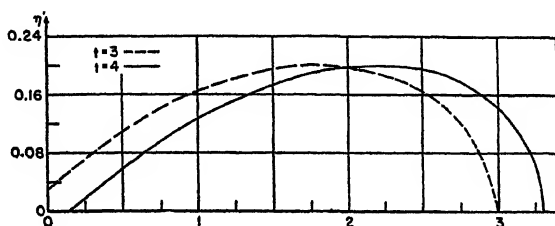


FIGURE 8

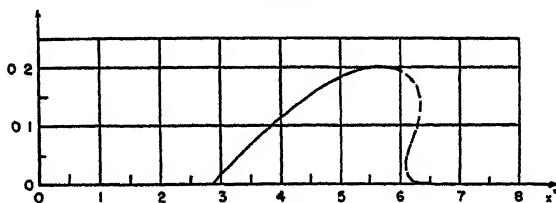


FIGURE 9

sion phase (*i.e.* A is negative in EQUATION 10). The third stage shown corresponds to the breaking time, while the fourth stage is again obtained by continuing the solution beyond this point. As one observes, the steepening of the wave front is very marked in this case, and one hardly doubts that the wave really would curl over and break very shortly after the point which we have somewhat arbitrarily defined as the breaking point.

In the case of water of uniform depth the mathematical solutions of the problems under discussion have an important property, which follows immediately from the interpretation of the solutions in terms of the characteristics: The maxima and minima of any pulse propagate unaltered in value into still water. In other words, *the crests and troughs propagate unaltered in height into still water of uniform depth.* Also, the results embodied in formulas 4 and 5 imply that a positive pulse or hump will always break eventually, no matter how small the ratio of amplitude to initial depth of the

water may be. This may account for some of the scatter in the observational data for waves on shallow sloping beaches given in a report of the

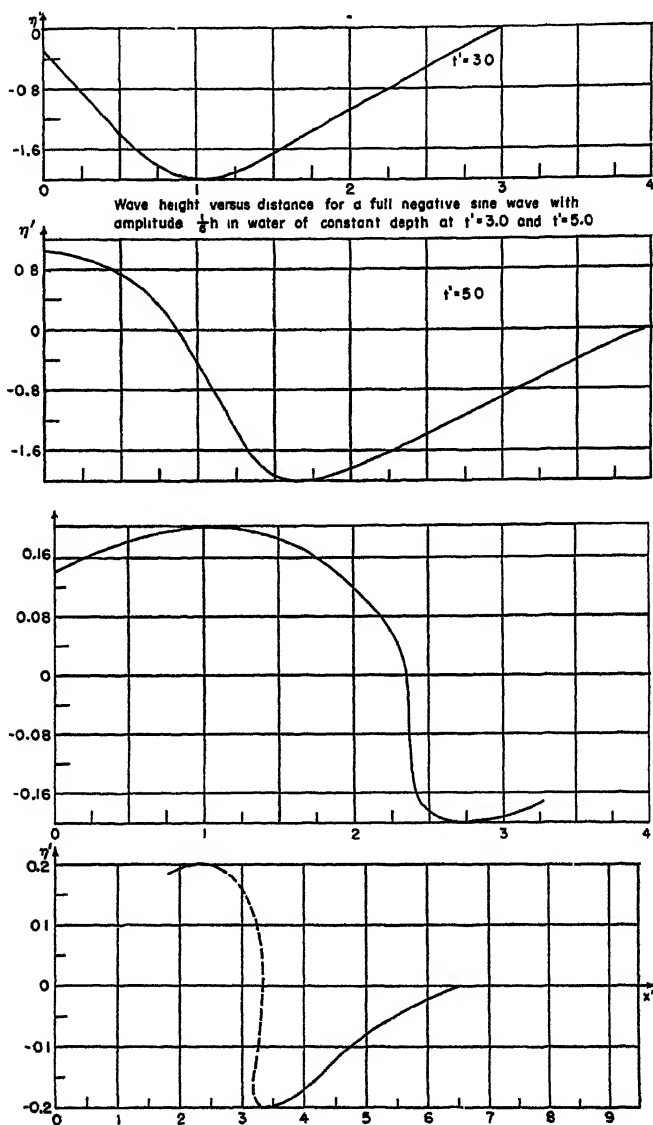


FIGURE 10

Hydrographic Office (Sverdrup and Munk, 1944), which show graphs of the ratio of breaker heights to deep water amplitudes and of still water depth, at the breaking point, to the amplitude in deep water as functions

of the "initial steepness" in deep water; the latter being defined as the ratio of amplitude to wave length in deep water. The "initial steepness" is, thus, essentially the quantity $\frac{1}{2}\omega$ in the formulas 4 and 5 and it should indeed be a significant parameter for the discussion of breaking phenomena. However, our above remarks indicate that the ratio of deep water amplitude to breaking height is not a good parameter for this purpose, since the breaking occurs for all values of this ratio in water of uniform depth. On a sloping beach, the circumstances are somewhat altered, since the wave amplitudes increase as the waves move toward shore. But it would seem rather likely that our general conclusion that the breaking is relatively insensitive to the ratio of amplitude to depth would still be valid, since the beach slopes are quite small as a rule.

It is of interest now to return to the problem with which we opened the discussion of the present section, *i.e.* to the problem of a tank with a movable end which is pushed into the water. As we have seen, the wave which arises will eventually break. Suppose now we assume that the end of the tank continues to move into the water with a uniform velocity. The end result after the initial curling over and breaking will be the creation of a steady progressing wave front which is very steep and turbulent behind which the water level is constant and the water has everywhere the constant velocity imparted to it by the end of the tank. Such a steady progressing wave with a steep front is called a bore. It is the exact analogue of a steady progressing shock wave in a gas. In FIGURE 11 we show a photograph, taken from *Probleme der Wasserwellen* (Thorade, 1931), of the bore which occurs in the T sien-Tang river as a result of the rising tide, which pushes the water into a narrowing estuary at the mouth of the river. The height of this bore apparently is as much as 20-30 feet. According to the theory presented above, this bore should have been preceded by an unsteady phase, during which the smooth tidal wave entering the estuary first curled over and broke.

We have, so far, used our basic theory to interpret the solutions of only one type of problem, *i.e.* the problem of the change of form of a pulse moving into still water of constant depth. The theory, however, can be used to study the propagation of a wave over a beach with decreasing depth just as well (Stoker, 1948), but the calculations are made much more difficult because of the fact that no family of straight characteristics exists unless the depth is constant. This problem, in fact, brings to the fore the difficulties of a computational nature which occur in important problems involving the propagation of flood waves and other surges in rivers and open channels in general, which are, however, still best treated by the method of characteristics.

On an actual beach, the motion of the water, of course, does not consist in the propagation of a single pulse into still water, but rather in the occurrence of an approximately periodic train of waves. However, observa-



tions from the air of the propagation speeds of the waves indicate that little or no reflection of the wave motion from the shore occurs. The incoming wave energy seems to be destroyed in turbulence due to breaking or to be converted into the energy of flow of the undertow. In other words, each wave propagates to a considerable degree, unaffected by the waves which preceded it. Another objection to our theory is the following: the degree to which the approximation furnished by our long wave-shallow water theory is accurate depends essentially on having curvatures of the wave surface which are small compared with the depth. But this condition is violated near the breaking point, hence our theory cannot be expected to be ac-

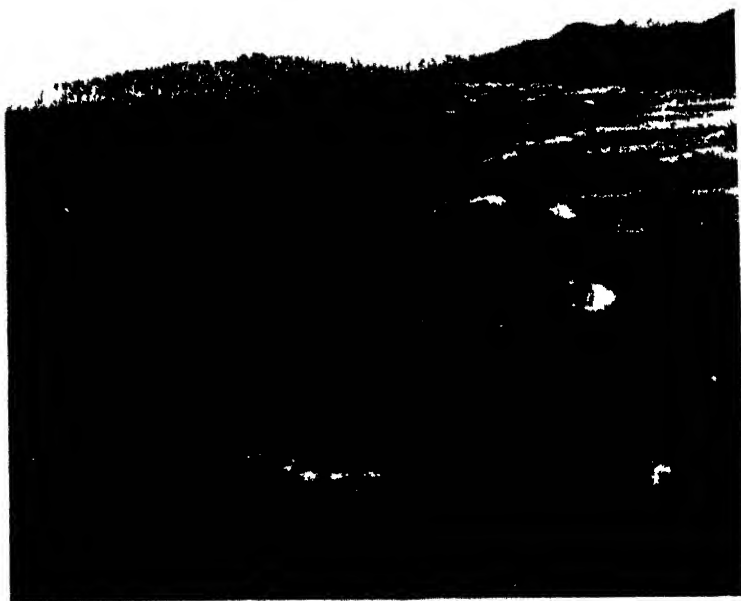


FIGURE 12

curate in this vicinity. Nevertheless, the theory should be valid, except near this point, in many cases of waves on sloping beaches, since the wave lengths are usually at least 10 to 20 times the depth of the water in the breaker zone, hence the theory presented above should certainly yield correct qualitative results and perhaps also reasonably accurate quantitative results.

It is, we repeat, essential for the applicability of the theory presented here that the depth should be small in comparison with the wave length, or, still better, that the curvature of the water surface should be small compared with the depth. If this condition is violated, the theory does not yield the breaking phenomena observed in nature. In FIGURE 12, we show a photo-

graph (given to the author by Walter Munk) of waves breaking in a fashion considerably at variance with the results of the theory presented here. We observe that the waves break, in this instance, by curling over slightly at the crest, but that the wave remains, as a whole, symmetrical in shape, while the theory presented in this paper yields a marked steepening of the wave front and a very unsymmetrical shape for the wave at breaking. In the case of FIGURE 12, the curvature of the wave surface would seem to be too large compared with the depth of the water to permit the use of the simple theory given here. When the curvature is small enough, however, the considerable amount of available experimental evidence furnished by the hydraulic engineers indicates that our theory presented here is accurate (cf. Favre, 1933, and Preiswerk, 1938).

Observation of cases like that shown in FIGURE 12 doubtless led to the formulation of the theory of breaking (Sverdrup and Munk, 1944) based on results taken from the study of what is called the solitary wave.* This is, by definition, a wave of finite amplitude consisting of a single elevation of such a shape that it can propagate unchanged in form. At first sight, this would seem to be a rather curious wave form to take as a basis for a discussion of the phenomena of breaking, since it is precisely the change in form resulting in breaking that is in question.

On the other hand, the waves often look as in FIGURE 12 and do retain, on the whole, a symmetrical shape,† with some breaking at the crest. Actually, the situation regarding the two different theories of breaking from the mathematical point of view is the following: Both theories are shallow water theories, *i.e.*, they are approximations to the exact hydrodynamical theory which are based on the assumption that the curvature of the water surface is small compared with the depth of the water. In fact, as Dr. Keller has indicated in his paper in the present volume, *the theory of the solitary wave can be obtained from the approximation of next higher order above that used in the present paper, if the assumption is made that the motion is a steady motion*. In other words, the theory used by Sverdrup and Munk is a shallow water theory that is more accurate for *steady waves* than the theory used by the author, which furnishes in principle the constant state as the only wave which can propagate unaltered in form. On the other hand, the theory presented here makes it possible to deal directly with the unsteady motions, while Sverdrup and Munk are forced to approximate these motions by a series of different steady motions. One could perhaps sum up the whole matter by saying that waves break in different ways depending upon the individual circumstances (in particular, the depth of the water compared with the wave length is very important), and the theory which should be used to describe the phenomena should be chosen accordingly.

* An interesting mathematical treatment of breaking phenomena from this point of view was given some time ago by Keulegan and Patterson (1940).

† Sverdrup and Munk, like the author, feel that, when considering breaking phenomena, each wave in a train can be treated with reasonable accuracy as though it were uninfluenced by the presence of the others.

Bibliography

- BOUSSINESQ, J. 1877. Essai sur la théorie des eaux courantes. Mémoires par divers savants 23: 24.
- BRUMAN, J. R. 1947. Application of the water channel-compressible gas analogy. A report to the U. S. Navy Department by North American Aviation, Inc.
- CORNISH, V. 1910. Waves of the Sea, and Other Water Waves. London.
- CORNISH, V. 1934. Ocean Waves and Kindred Physical Phenomena. Cambridge.
- COURANT, R., & K. O. FRIEDRICHS. 1944. Supersonic flow and shock waves. Report prepared for the Applied Mathematics Panel of the National Defense Research Committee. This report, in revised and extended form, is to be issued as a book.
- EINSTEIN, H. A., & E. G. BAIRD. 1946. Progress report of the analogy between surface shock waves on liquids and shocks in compressible gases. Report prepared for the Bureau of Ordnance of the U. S. Navy Department by the Hydrodynamics Laboratory, California Institute of Technology.
- FAVRE, H. 1933. Contribution à l'étude des courants liquides. Publication du laboratoire de recherches hydrauliques annexé à l'école polytechnique fédéral de Zurich.
- VON KÁRMÁN, T. 1938. Eine praktische anwendung der analogie zwischen überschallströmung in gasen und überkritischer strömung in offenen gerinnen. Zentralblatt für angewandte Mathematik & Mechanik 18 (1).
- KEULEGAN, G. H., & G. W. PATTERSON. 1940. Mathematical theory of irrotational translation waves. J. Res. Nat. Bureau Stand. U. S. Department of Commerce: 47.
- LAGRANGE, I. L. 1781. Mémoire sur la théorie des mouvements des fluides. Oeuvres IV.
- LAMB, W. 1945. Hydrodynamics. American Edition.
- PREISWERK, E. 1938. Anwendung gasdynamischer Methoden auf Wasserströmungen mit freier Oberfläche. Mitteilungen aus dem Institut für Aerodynamik, Technische Hochschule, Zürich.
- RIABOUCHINSKY, D. 1932. Sur l'analogie hydraulique des mouvements d'un fluide compressible. C. R. Acad. Sci. Paris. 195: 998.
- STOKER, J. J. 1948. Formation of breakers and bores. Communications on Applied Mathematics I (1): 1-87.
- SVERDRUP, H., & W. MUNK. 1944. Breakers and surf. Hydrographic Office, U. S. Navy Department.
- SVERDRUP, H., & W. MUNK. 1947. Wind, sea, and swell: Theory of relations for forecasting. H. O. Navy Department.
- THORADE, H. F. 1931. Probleme der Wasserwellen. Probleme der kosmischen Physik, Bd. 13-14.

THE SOLITARY WAVE THEORY AND ITS APPLICATION TO SURF PROBLEMS*

BY WALTER H. MUNK

Scripps Institution of Oceanography and Institute of Geophysics, University of California

Introduction

The purposes of this paper are: (a) to give a summary of useful relations derived by means of the solitary wave theory, and to plot these relations using dimensionless parameters for the purpose of making the theory accessible to numerical examples;† (b) to review various studies at the Scripps Institution dealing with the application of this theory to surf problems; and (c) to discuss the problem of sand transport in or near the surf zone, in the light of the solitary wave theory.

This investigation represents part of a general project undertaken during the war for the purpose of providing useful wave forecasts for the amphibious forces. By 1943, methods for forecasting sea and swell had been developed^{1,2} and a study of the transformation of waves in shallow water was initiated for the purpose of extending the wave forecasts right into the surf zone. It should be noted that the outer edge of the surf zone (the greatest depth where waves break) is usually the most critical from the point of view of bringing landing craft ashore.

The problem was attacked in three ways: (a) by field observations along the East Coast by the Woods Hole Oceanographic Institution and along the West Coast by the Scripps Institution of Oceanography; (b) by laboratory observations at the Beach Erosion Board wave tank, in Washington, D. C., and later at the Department of Engineering of the University of California in Berkeley, California; (c) by theoretical studies.

A theoretical investigation by Burnside,³ based on the assumptions of constancy of wave periods, conservation of energy, and the linear shallow water (Airy) wave theory, reveals that the waves decrease somewhat in height after entering shallow water, reach a minimum height and then increase.^{4, 5} The *initial decrease* in wave height had been noticed by O'Brien in laboratory investigations. A comparison between the subsequent *increase in height* as derived from Burnside's equations with that obtained from field and laboratory observations mentioned above, showed the computed increase to be considerably smaller than the observed increase. This discrepancy became increasingly large the nearer one came to the breaking zone, the zone most important for practical forecasts.

One reason for this discrepancy is contained in an assumption underlying the linear Airy theory, namely that the wave height be small compared to

* Contribution from the Scripps Institution of Oceanography, New Series No. 406. This work represents results of research carried out for the Hydrographic Office, the Office of Naval Research, and the Bureau of Ships of the Navy Department under contract with the University of California.

† PLATES 1-12 at end of the paper

the water depth. This assumption is not fulfilled in the vicinity of the breaker zone. Another theoretical approach was therefore undertaken, again based on the assumptions of constancy in period and conservation of energy, but this time making use of Stokes's theory for oscillatory waves of *finite height* in shallow water. Although this approach led to an adjustment in the theoretical results in the proper direction, it did not yield useful results because the theory involves infinite series which converge more and more slowly as one approaches the breaker point.

Thus, the assumptions in Airy's theory do not hold, and Stokes's theory becomes unmanageable, both shortcomings becoming increasingly serious the closer one approaches the zone of greatest practical interest, the breaker zone. Faced with this situation and an immediate need for a method of forecasting breaker heights, a manual which was based principally on empirical relationships was prepared.⁶

The reasons for the shortcomings of Stokes's theory became apparent soon after the publication of the forecasting manual. As waves travel into water of depth less than, say, three times the wave height, the previously flat crests "hump" into narrow crests separated by long flat troughs, and the character of these isolated crests scarcely depends upon the distance L between the crests (FIGURE 1). Yet the wave length L is contained in the two fundamental parameters appearing in the series of Stokes's theory: the *relative depth* (depth/wave length, h/L) and the *wave steepness* (wave height/wave length, H/L). The solitary wave theory on the other hand, contains a fundamental parameter that is independent of wave length: the *relative wave height* (wave height/depth, H/h). The application of the solitary wave theory* was suggested also by an obvious resemblance between the theoretically derived wave profile and the observed profile in the region just outside the breaker zone.

A third theoretical approach based on the assumptions of constancy of wave period, conservation of energy, and the solitary wave theory immediately led to agreement with earlier observations, and an improvement of the forecasting graphs.^{1, 7} These modified graphs are still being used for forecasting breaker characteristics. The confidence gained as a result of the agreement between theory and observation led to the curtailment of an ambitious program of observations. The theory has since been applied to the following problems: humping of waves just before breaking, wave refraction, prediction of breaker height and depth of breaking, and the prediction of longshore currents. It has also been used, so far without confirmation, to compute the forces exerted by waves just prior to breaking.

Boussinesq developed the first well-founded theory of solitary waves in 1871 to explain the propagation of *isolated* elevations and depressions which had been observed by Scott-Russell in 1844 during experiments in straight

* It is of interest to note that Stokes's theory and the solitary wave theory represent the two limiting cases of a more general type of periodic waves.⁸

channels with rectangular cross-sections. Lord Rayleigh in 1876, and McCowan^{11, 12} in 1891, carried Boussinesq's solution to a higher degree of approximation. Thereafter, the solitary wave theory received little attention, until Keulegan and Patterson⁹ published their exhaustive treatise about fifty years later.

In theory, a solitary wave consists of a single crest of infinite length, and the application to surf problems represents therefore a departure from the type of phenomenon for which the theory was intended. Yet, because the energy of the solitary waves is largely confined within a relatively narrow band about the crest, it seems reasonable to apply the theory to periodic humping crests of ocean swell. Indeed, it has been shown that the solitary wave represents an extreme case for certain types of periodic waves.⁸

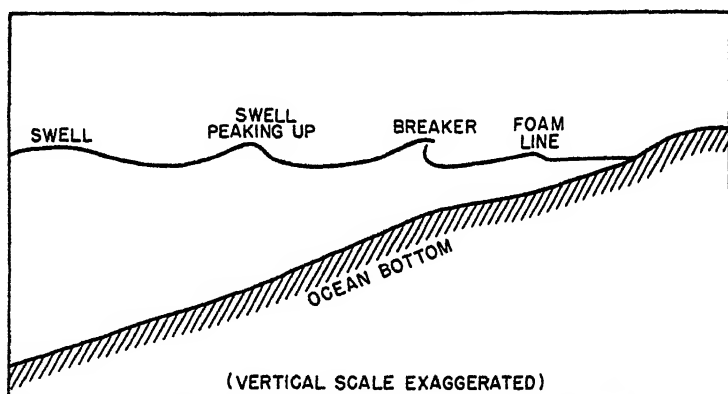


FIGURE 1. Schematic presentation of change in wave shape as wave advances into shallow water

Theory of Solitary Waves

Significant Non-Dimensional Parameters. The notation is explained in PLATE 1. Whenever possible non-dimensional parameters are denoted by capital letters. Linear dimensions are divided by the depth h to give the non-dimensional coordinates

$$X = x/h, \quad Z = z/h, \quad (1)$$

the relative elevation of the free surface

$$\mu = \eta/h, \quad (2)$$

and the relative wave height

$$\gamma = H/h. \quad (3)$$

At the bottom $Z = 0$, at the crest of the wave $Z = 1 + \gamma$. The non-dimensional components of horizontal and vertical orbital velocity are

$$U = u/C, \quad W = w/C, \quad (4)$$

where C is the wave velocity.

In this report, various characteristics of the solitary waves will be expressed as functions of the distance coordinate, X , measured relative to the position of the (moving) wave crest. Sometimes, it is more convenient to consider changes with time at a given point. If $t = 0$ denotes the time of crest passage, then $x = Ct$, $X = (C/h)t$, and the coordinate X may be considered a non-dimensional time coordinate, from which the actual time follows according to the equation

$$t = \frac{h}{C} X. \quad (5)$$

The First Approximation

Boussinesq's approach. Boussinesq^{9*} has obtained a solution to the equations of motion for irrotational, non-divergent flow, subject to the usual surface and bottom boundary conditions. The solution is based on the expansion of the velocity potential in a power series

$$\phi = \sum_{n=0}^{\infty} \phi_n z^n$$

and applying the method of successive approximations.

Wave velocity and profile. Retaining only the first two terms in the power series, Boussinesq obtains for the wave velocity

$$C = \sqrt{g(h+H)} = \sqrt{gh(1+\gamma)} \quad (6)$$

and for the wave profile

$$\mu = \gamma \operatorname{sech}^2 \left(\sqrt{\frac{3\gamma}{4}} X \right). \quad (7)$$

PLATE 2 gives the wave velocity as a function of depth and wave height.

Volume and mass transport. The volume per unit crest length contained above $Z = 1$ between $-X$ and $+X$ is given by

$$Q' = 2 \int_0^x \eta \, dx = 2h^2 \int_0^X \mu \, dX = 4h^2 \sqrt{\frac{\gamma}{3}} \sqrt{1 + \frac{\mu}{\gamma}}. \quad (8)$$

Let Q designate the value of Q' for $X = \pm \infty$. Then

$$Q = 2 \int_0^{\infty} \eta \, dx = 2h^2 \int_0^{\infty} \mu \, dX = 4h^2 \sqrt{\frac{\gamma}{3}} \quad (9)$$

is the total volume of a solitary wave above the still water level. PLATE 3 shows the partial volume Q'/Q as a function of X for various values of γ . Taking for example $\gamma = 0.5$, one finds that 90 per cent of the volume of a solitary wave is contained between $X = \pm 2.4$, and 98 per cent between $X = \pm 3.8$.

* Equations which are derived in this publication will be given here without derivation.

This concentration near the crest is the basis for applying the expressions for the total volume of a solitary wave to oscillatory waves of finite length.* Since a volume transport equal to Q takes place essentially during a period, T , of the oscillatory waves, the mean transport per unit time equals Q/T , and the volume transport velocity averaged from surface to bottom equals

$$\bar{v} = \frac{Q}{hT}. \quad (10)$$

Substituting from EQUATION (9) gives

$$\bar{v} = \frac{4h}{T} \sqrt{\frac{\gamma}{3}}. \quad (11)$$

Energy. The energy consists approximately of equal parts of potential and kinetic energy. The energy per unit crest width between $-X$ and $+X$ equals

$$E' = \frac{8}{3} \rho g h^3 \gamma \left(2 + \frac{\mu}{\gamma} \right) \sqrt{\frac{\gamma}{3}} \sqrt{1 - \frac{\mu}{\gamma}} \quad (12)$$

and the total energy, between $-\infty$ and $+\infty$,

$$E = \frac{8}{3} \rho g h^3 \gamma \sqrt{\frac{\gamma}{3}}. \quad (13)$$

According to PLATE 3 for $\gamma = 0.5$, 90 per cent of the energy is contained between $X = \pm 1.6$, and 98 per cent between $X = \pm 2.1$.

Particle Displacements and Trajectories. The particles are sensibly at rest until $X \doteq 10$. They then move forward and upward and attain their maximum velocity at the instant the crest passes. Eventually, they reach a new rest position, having been displaced from their previous rest position in the direction of wave motion a horizontal distance r from their original position.

The foregoing description applies to the trajectory in still water. We shall deal later with the case of waves traveling toward a beach, where the particle motion will be modified by a return flow.

Let r' be the horizontal displacement of a particle at a time t relative to its position at $t = 0$ (the instant of crest passage). Let u be the horizontal component of particle velocity, and let x denote the horizontal distance between the particle and the wave crest. Then $x = 0$ for $t = 0$. At a time dt the crest is at $C dt$, the particle at $u dt$ and $dx = (C - u) dt$. Thus

$$r' = \int_0^t u dt = \int_0^x \frac{u}{C - u} dx. \quad (14)$$

The horizontal velocity and displacement can be replaced by their respective mean values \bar{u} and \bar{r} , the average being formed between surface and

* See EQUATIONS (38) and (39) for the approximation involved.

bottom, since $|(u - \bar{u})| \ll \bar{u}$, and $|(r - \bar{r})| \ll \bar{r}$. It follows from the equation of continuity^{10*} that

$$\frac{\bar{u}}{\bar{C}} = \frac{\eta}{h + \eta}, \quad (15)$$

so that

$$\bar{r}' = \int_0^\infty \frac{\eta(h + \eta)}{1 - \eta/(h + \eta)} dx = \frac{1}{h} \int_0^\infty \eta dx = \frac{1}{2} \frac{Q'}{h} \quad (16)$$

and

$$\bar{r} = \frac{2}{h} \int_0^\infty \eta dx = \frac{Q}{h}. \quad (17)$$

The latter formula follows also directly, since the total volume of water, Q , transported by a wave must equal the mean displacement \bar{r} times the depth h .

The corresponding non-dimensional displacements are

$$\bar{R}' = \frac{\bar{r}'}{h} = \frac{Q'}{2h^2} = 2 \sqrt{\frac{\gamma}{3}} \sqrt{1 - \frac{\mu}{\gamma}} \quad (18)$$

and

$$\bar{R} = \frac{\bar{r}}{h} = \frac{Q}{h^2} = 4 \sqrt{\frac{\gamma}{3}}. \quad (19)$$

The ratios $2\bar{r}'/\bar{r} = 2\bar{R}'/\bar{R} = Q'/Q$ are plotted as functions of X in PLATE 3.

The horizontal displacement and velocity have been assumed uniform from top to bottom. To the same degree of approximation it follows from the equation of continuity that the vertical displacement s and velocity w must be proportional to z_0 , the *initial* elevation of the particle. Let s' denote the instantaneous vertical *displacement* of a particle above its *initial* position. For a particle initially at the surface, $s' = \eta$; for any other particle, $s' = (z_0/h)\eta$, or, letting $S' = s'/h$, $Z_0 = z_0/h$,

$$S' = Z_0 \mu. \quad (20)$$

Eliminating μ between (18) and (20) gives

$$S' = Z_0 [\gamma - \frac{3}{4}(\bar{R}')^2] \quad (21)$$

for the particle trajectory of a solitary wave. The orbits of *surface* particles are drawn in PLATE 4 for various values of γ . The orbital motion at any other depth can be found by multiplying the vertical coordinate by Z_0 . The dashed lines give the positions occupied by the particles at various times. As the wave first approaches, each particle begins to move forward from rest at an inclination

$$\frac{dS'}{d\bar{R}'} = -Z_0 \sqrt{3\gamma}. \quad (22)$$

* Also p. 71 of reference 9.

Its velocity increases until, directly beneath the crest, it will move horizontally at a distance $Z_0\gamma$ above its original position. The particle then moves downward again, slows up, and reaches its original elevation moving at an angle equal and opposite to (22).

For small values of γ , EQUATION 21 takes the form

$$(\bar{R}')^2 = -\frac{4}{3Z_0} S' \quad (23)$$

which is that of a *parabola* with the origin at the vertex, and the focus at $-1/3Z_0$.

Higher Approximations

McCowan's approach. McCowan¹¹ reduces the problem to one of steady motion by impressing upon the fluid a uniform current of velocity $-C$; by assuming the velocity potential, ϕ , and the stream function, ψ , to satisfy the complex equation

$$\frac{1}{h} \frac{\psi + i\phi}{C} = -(Z + iX) - \frac{N}{M} \tan \left[\frac{1}{2} M(Z + iX) \right] \quad (24)$$

the assumptions of non-divergent and irrotational flow are satisfied without approximation. Substituting into the equations of motion and satisfying the boundary conditions, leads to expressions for C , μ , Q , E , of which the preceding solutions by Boussinesq are a first approximation.

It will be shown that, for most applications, the first approximation is adequate. Only in problems involving an accurate knowledge of the orbital velocity Boussinesq's treatment is not sufficiently accurate. This lack of accuracy is more pronounced the closer one approaches the extreme form of the breaking wave.

Orbital velocities. Using non-dimensional parameters, McCowan's equations for the horizontal and vertical components of orbital velocity become, in our notation

$$U = N \frac{1 + \cos MZ \cosh MX}{(\cos MZ + \cosh MX)^2}, \quad (25)$$

$$W = N \frac{\sin MZ \sinh MX}{(\cos MZ + \cosh MX)^2}, \quad (26)$$

where the parameters M and N are functions of γ (PLATE 5) and are defined by the equations

$$\gamma = \frac{N}{M} \tan^{1/2} [M(1 + \gamma)], \quad (27)$$

$$N = \frac{2}{3} \sin^2 \left[M \left(1 + \frac{2}{3} \gamma \right) \right]. \quad (28)$$

As a first approximation, $M = \sqrt{3\gamma}$, $N = 2\gamma$. The ratio $2\pi/M$ may be interpreted as an effective non-dimensional wave length, since 92 per cent of the volume, and 99 per cent of the energy of a solitary wave are contained within the zone $X = \pm 2\pi/M$. It will be shown that $N = RM = 2\pi R/L_{\text{eff}}$, so that N equals 2π times the horizontal displacement of a particle, divided by the effective wave length.

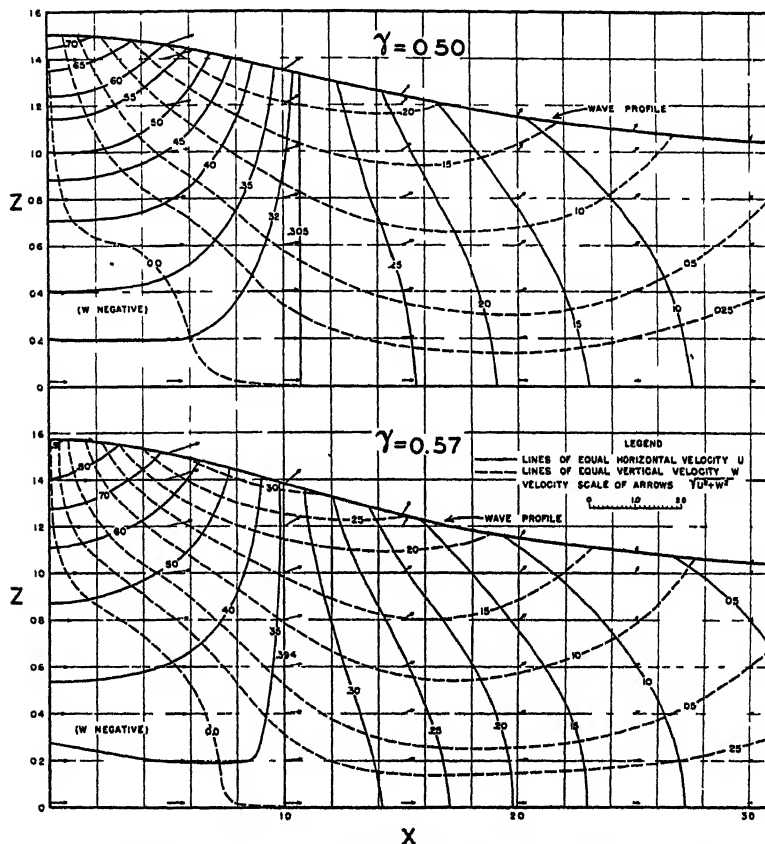


FIGURE 2. Orbital motion in a solitary wave according to Boussinesq. The presentation is analogous to PLATE 8.

PLATES 6-9 show the orbital velocities as functions of X and Z for eight selected values of γ . The surface profiles are plotted according to McCowan's equation

$$\mu = \frac{N}{M} \frac{\sin M(1 + \mu)}{\cos M(1 + \mu) + \cosh MX} \quad (29)$$

to which EQUATION 7 is a first approximation. FIGURE 2 is analogous to

PLATE 8, except that it is based on Boussinesq's theory. The condition that families of U and W be orthogonal is not even approximately fulfilled in certain regions of FIGURE 2, whereas this condition is exactly fulfilled in PLATE 8.

Directly beneath the crest $X = 0$, $W = 0$, and

$$U = N \frac{1 + \cos MZ}{\sin^2 MZ} \quad (30)$$

as shown in PLATE 10. Along the bottom $Z = 0$, $W = 0$, and

$$U = N \frac{1}{1 + \cosh MX} \quad (31)$$

as shown in PLATE 11. PLATES 5-11 permit rapid calculation of the orbital velocities for any chosen set of conditions.

McCowan's solution for extreme crest angle. For the special case of a breaking wave McCowan¹² has modified his original treatment. Assuming as a criterion for breaking that $U = C$ at the very crest, McCowan finds the crest to be formed by two branches equally inclined to the bottom and cutting at an angle of 120° . The corresponding value of the relative wave height is

$$\gamma_b = \frac{1}{2} \tan (1 \text{ radian}) = 0.7813 \dots \quad (32)$$

where the b denotes *breaker*. In the neighborhood of the crest the wave profile is given by

$$\mu = 1.04 e^{-x} - 0.44 e^{-2x} + \dots \quad (33)$$

and the orbital velocities by

$$-(U + iW) \doteq 0.80 \sqrt{Z^* - i\bar{X}} [1 + 0.084(Z^* - i\bar{X})] \quad (34)$$

where

$$Z^* = Z - (1 + \gamma) \quad (35)$$

is the non-dimensional vertical coordinate relative to the elevation of the crest. In the preceding plates, the profile and orbital velocities for the case $\gamma = 0.78$ were computed according to EQUATIONS 32 and 33.

Particle displacement and drift. It has already been stated that the horizontal forward displacement of particles in a solitary wave is almost uniform from top to bottom, so that

$$|(R - \bar{R})| < \bar{R}, \quad (36)$$

where \bar{R} is the mean displacement. Boussinesq's expression for the mean displacement is given by EQUATION 19.

When the theory is carried to a high degree of approximation, one finds

$R - \bar{R}$ to be positive near the surface, negative near the bottom, *i.e.*, the particles are farther displaced in the direction of wave travel near the surface than near the bottom. This feature is of considerable interest in the study of sand movements, for should the total shoreward transport be balanced by a uniform return flow, the combined flow pattern would exhibit a net *drift*

$$V - \bar{V} = \frac{R - \bar{R}}{T} \quad (37)$$

which is directed shoreward (+) near the surface, seaward (-) near the bottom.

The calculation of $R(Z)$ involves successive approximations and is very laborious. The method is given in APPENDIX 1. PLATE 12 gives the results of these calculations for various values of γ . The special case of breaking $\gamma = 0.78$ has again been treated by McCowan's modified theory. Surface displacements ($R_1 - \bar{R}$) and bottom displacements ($R_0 - \bar{R}$) are shown in the inset. Numerical values in PLATE 12 may be in error by as much as 10 per cent.

Applications of the First Approximation

Restrictions in the application of the theory. The solitary wave theory involves the assumptions of an infinitely long wave, and of no boundary in the direction of wave travel. These are clearly not fulfilled in the case of ocean waves traveling into shallow water. The presence of the beach imposes an additional boundary which leads to the establishment of a return flow system that is superimposed upon the still water orbital motion. Since the velocity of this flow pattern is small compared to the wave velocity, it may be assumed that the equations for the *velocity, profile, volume* and *energy* are little altered. On the other hand, the effect of the return flow pattern on the *orbital motion* is not at all negligible. Applications of the theory which do not involve orbital motion can be treated by the approximate solutions derived by Boussinesq.

The assumption of a single solitary wave is fulfilled to a high degree of accuracy if the *actual* wave length L of the waves exceeds the *effective** wave length of $2\pi/M$ or, (since $L \doteq T\sqrt{gh}$), if the wave period T exceeds T_{eff} , where

$$T_{\text{eff}} = \frac{2\pi}{M} \sqrt{\frac{h}{g}} \quad (38)$$

as given by Bagnold.¹³ For breaking waves

$$(T_{\text{eff}})_b = 1.17 \sqrt{h_b} = 1.32 \sqrt{H_b}$$

* See discussion following EQUATION 28.

where T is in seconds, h_b and H_b in feet. PLATE 2 gives T_{eff} for various values of H and h .

In a true solitary wave of infinite length the mean depth coincides with the depth at $\pm \infty$. In the case of periodic waves the mean depth will exceed the depth at the trough by an amount $Q/L = Q/CT$, or in view of (6) and (9)

$$\bar{h} = h \left[1 + \frac{4}{\sqrt{g}} \frac{1}{T} \sqrt{\frac{H}{3(1 + \gamma)}} \right]. \quad (39)$$

In most subsequent applications this correction has been neglected.

The height of breaking. The derivation of breaker heights is based on the assumptions that the period remains constant and that energy is conserved.⁴ Neglecting for the present the effect of refraction due to waves approaching the beach at an angle, the rate of energy flow in deep water (subscript "0" indicates deep water values) per unit crest width

$$\frac{C_0}{2} \frac{E_0}{L_0} = \frac{C_0}{2} \left(\frac{1}{8} \rho g H_0^2 \right) \quad (40)$$

must equal the energy flow at the breaking point,

$$C_b \frac{E_b}{L_b}, \quad (41)$$

where E_b/L_b is the mean energy per unit surface area. Since the wave period is assumed constant,

$$T = \frac{L}{C} = \frac{L_b}{C_b}. \quad (42)$$

Equating (40) and (41) and using (42), (13) and (32) gives

$$\frac{H_b}{H_0} = \frac{1}{3.3 \sqrt[3]{H_0/L_0}}. \quad (43)$$

This relationship is shown by the line on the left side of FIGURE 3, where the limitations imposed by (38) can be expected to apply. The line to the right marked "Airy Wave Theory" gives the change in wave height according to EQUATION 18 of "Breakers and Surf".⁶ In the derivation of this equation it is assumed that no change in the *shape* of the wave takes place. Therefore the Airy theory is more likely to apply to waves of large initial steepness, which break shortly after entering shallow water without suffering much change in shape.

With the relationship fixed by theory for both large and small values of steepness, there is little leeway in drawing a continuous curve to represent the relationship between H_b/H_0 and H_0/L_0 . In FIGURE 3 the theoretical relationships have been taken as valid for values of H_0/L_0 less than .006 and greater than .06. In the mid-range of wave steepness, that is for H_0/L_0

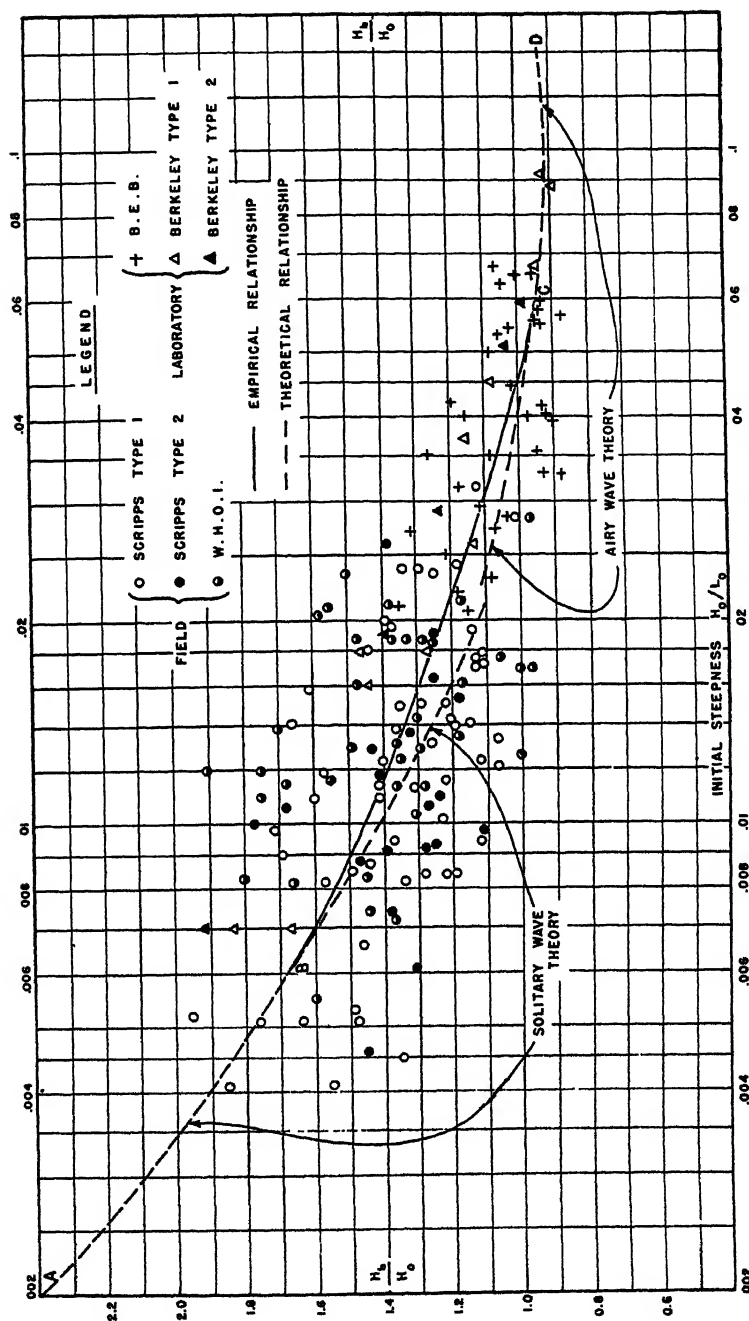


FIGURE 3. Comparison between predicted and observed breaker heights.

between .006 and .06, the solid line BC is taken to represent the relationship between H_b/H_0 and H_0/L_0 . The line $ABCD$ together with refraction considerations forms the basis of the present method for predicting breaker heights.

Measurements taken at various localities are denoted by different symbols, as shown in the legend of FIGURE 3. All values have been tabulated in APPENDIX 2. The open and filled circles give ratios of wave heights obtained from photographs taken against the Scripps Institution pier¹⁴ and will be referred to as the "Scripps Leica Data". The wave height in deep water is found from measurements at the end of the pier and EQUATION 18, "Breakers and Surf".⁶ The filled circles marked "Type 2" in the legend refer to waves which are breaking behind a bar in water of increasing depth.¹⁵ All other conditions are included under Type 1. This grouping does not introduce a consistent error with regard to H_b/H_0 although it will later be shown that it

TABLE 1
EFFECT OF BEACH CHARACTERISTICS ON BREAKER HEIGHT

Source	Beach slope or type	Number of Measurements	$\frac{H_b/H_0}{\bar{H}_b/\bar{H}_0}$
Woods Hole measurements		38	1.104
Scripps Pier Leica data	Type 1	55	0.96
	Type 2	19	0.96
Beach erosion board wave tank	0.159	13	0.92
	0.049	15	1.08
	0.030	9	0.91
Berkeley wave tank	0.072	6	1.11
	0.054	5	1.01
	0.009	5	1.10

has important bearing upon the depth of breaking. The half-filled circles refer to measurements made along South Beach, Martha's Vineyard, Massachusetts.¹⁶ It is not known whether these measurements represent Type 1 or Type 2 in the foregoing sense. The crosses refer to measurements at the Beach Erosion Board Laboratory in Washington.¹⁷ These measurements show good agreement with theory. The triangles are based upon measurements in the wave tank at Berkeley¹⁸ and include some measurements of waves breaking on a flat beach immediately following a very steep beach.

To investigate the causes for the large scatter in FIGURE 3, the value \bar{H}_b/\bar{H}_0 corresponding to the observed steepness H_0/L_0 was read off the curve for each individual point, and the ratio $(H_b/H_0)/(\bar{H}_b/\bar{H}_0)$ formed. Average values of $(H_b/H_0)/(\bar{H}_b/\bar{H}_0)$ for beaches of Types 1 and 2 and for various bottom slopes in the laboratory are summarized in TABLE 1. This ratio appears to be independent of beach type, and not to vary consistently with bottom slope.

In order to find the possible role played by friction, average values of $(H_b/H_0)/(\bar{H}_b/\bar{H}_0)$ have been formed for various ranges in s/L_0 where s is the distance the wave had traveled from the time it enters "very shallow water"¹⁸ ($h/L_0 < 0.05$) until the time it breaks. The ratio s/L_0 gives a measure of the extent to which waves are subject to relatively large frictional dissipation. TABLE 2 indicates a small though consistent effect of friction, amounting to a reduction in wave height of 10 per cent at the very most. It may be concluded that at least some of the scatter of points in FIGURE 3 can be ascribed to friction. Except over very gently sloping beaches, the effect of friction is likely to be small, and for practical forecasting of breaker heights the assumption of conservation of energy appears justified.

The depth of breaking. The data pertaining to the depth of breaking are even more scattered than the data for the height of breaking. In the laboratory, appreciable errors in observation are introduced by the effect of

TABLE 2
EFFECT OF BOTTOM FRICTION ON BREAKER HEIGHT

$\frac{s}{L_0}$.00-.29	.30-.59	.60-.89	.90-1.19	1.20-1.99
Number of observations	13	25	33	32	7
$\frac{H_b/H_0}{\bar{H}_b/\bar{H}_0}$	1.04	1.02	0.99	0.91	0.90

"surging," and in the field changes, in the bottom slope, such as those due to sand bars, become an important factor. The reason for the latter feature is probably that the depth of breaking is a *kinematic* problem, involving particle velocities, and subject to the boundary conditions inflicted by the bottom. On the other hand, the breaker height depends chiefly on energy considerations and is not *directly* dependent upon any bottom boundary conditions.

According to EQUATION 32, waves may be expected to break in water of depth 1.28 times the breaker height. There have been frequent reports of waves breaking in much deeper water. Regarding these observations Gaillard¹⁹ writes

"Prof. G. B. Airy has shown mathematically that when the depth is variable it is impossible that a series of waves can exist having oscillatory motion of the particles and satisfying the equations of continuity and of equal pressure. While the continuity holds equal pressure *will* exist; therefore in such a case continuity *must* cease, i.e., the water becomes broken. This is, in his opinion, the explanation of the fact that the sea in places breaks over deeply submerged banks, shoals, or reefs, where the waves in general are high."

Thus, in the case of sudden changes in depth, the ratio h_b/H_b may become very large, but in this report only waves breaking on sloping bottoms will be considered.

TABLE 3 gives average values of h_b/H_b for various field and laboratory investigations. Observations of waves breaking *behind* a sand bar in water of *increasing* depth (Type 2) have been eliminated from the field data. Tank observations during which the waves were *generated* in shallow water ($h/L_0 < 0.5$) are likewise eliminated.²⁰

Values marked "S.I.O. surf" are based on a small number of precise *simultaneous* measurements of h_b and H_b . The depth of breaking was measured by carrying a hollow tube into the surf zone and measuring the mean water level from a marker attached to a float inside the tube. The lower end of the tube was pushed a short distance into the sand in order to reduce the up and down movement of the water inside the tube, making an accurate determination of the still water depth possible. The breaker height was measured at the same time from markings on the outside of the tube. Three

TABLE 3
SUMMARY OF OBSERVATIONS OF DEPTH OF BREAKING

<i>Location</i>	<i>Number of observations which were averaged</i>	<i>Mean wave steepness, H_0/L_0</i>	h_b/H_b
Estero Bay	244	.005	1.30
SIO Daily	238	.008	1.30
SIO Leica	56	.013	1.21
SIO surf	28	—	1.34
WHOI	17	.014	1.10
Berkeley Tank	2	.024	1.28
BEB Tank	27	.040	1.30
Lake Superior	134	.070	1.56
	746	Weighted Mean:	1.34

observers stationed in the surf zone took simultaneous readings of the water depth, the elevation of the trough and of the breaking crest. Unless the wave broke right at the tube, the measurements were eliminated. Results are shown in FIGURE 4. The solid line gives the relationship of $h_b/H_b = 1.28$, and is in fair agreement with the observations.

On the whole, the observations support the solitary wave theory. This conclusion is also borne out by a number of qualitative statements. According to a British report by the Superintendent of Mine Design: "For a wave of any period and height the waves break when the depth of water is 4/3 of the wave height." In a Canadian Intelligence report for the Mediterranean area a range in values from 1.2 to 1.45 was quoted. A report by the Swell Forecast Section of the British Admiralty gave an average value of 1.5. In a study of refraction at Monterey, the depth of breaking was obtained from aerial photographs, and the height of breaking from measurements of wave heights in deep water and from refraction diagrams. An average of

$h_b/H_b = 1.14$ was obtained from seven measurements, but due to the uncertainties involved, not much emphasis should be placed on this value.

Peaking of waves just outside breaker zone. Just outside the surf zone the wave crests rise sharply from the water surface (FIGURE 1). Even over gently sloping beaches the transition from the region of flat symmetrical waves to the region of humped wave crests is remarkably sudden.

The symmetrical wave profiles conform to the Airy's theory; the humped

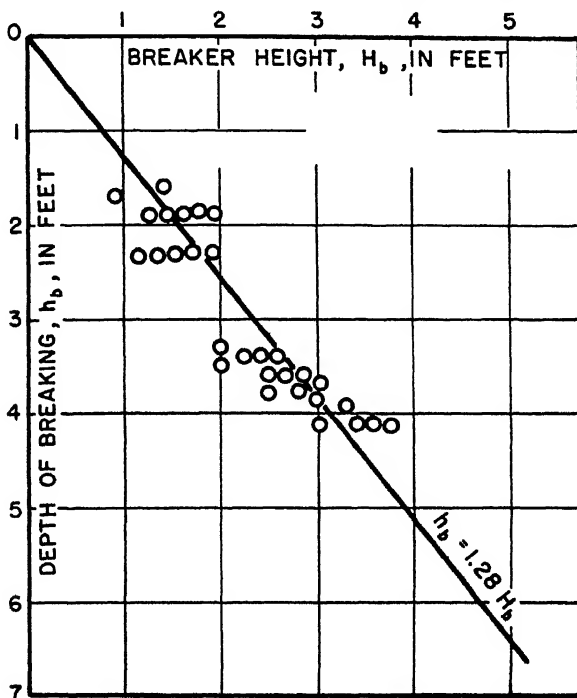


FIGURE 4. The circles denote individual observations of H_b and h_b conducted in the surf zone off the Scripps Institution (see text). Observations are in fair agreement with the theoretical relationship denoted by the line. (*Scripps Beach Measurements.*)

wave profiles conform to the solitary wave regime. In order to obtain some criterion for the outer boundary of the solitary wave regime, we shall compute changes of wave height with depth, according to both theories, and compare the computed values with observations taken along the Scripps Institution pier.

Assume again that the wave period remains constant and that energy is conserved. Since we are dealing with very shallow water where the energy is propagated at wave velocity the latter condition requires that

$$C_1 E_1 = C_2 E_2, \quad (44)$$

the subscripts referring to two given locations.

For an Airy wave $E \sim H^2$, $C \sim h^{1/2}$, so that

$$\frac{H_2}{H_1} = \left(\frac{h_1}{h_2} \right)^{1/4}. \quad (45)$$

For a solitary wave $C^2 = gh(1 + \gamma)$ (EQUATION 6), $E = (8/3)\rho gh^2\gamma\sqrt{\gamma/3}$ (EQUATION 13), and, according to (44)

$$\frac{H_2}{H_1} = \left(\frac{h_1}{h_2} \right)^{4/3} \left(\frac{1 + \gamma_1}{1 + \gamma_2} \right)^{1/3} \doteq \left(\frac{h_1}{h_2} \right)^{4/3}. \quad (46)$$

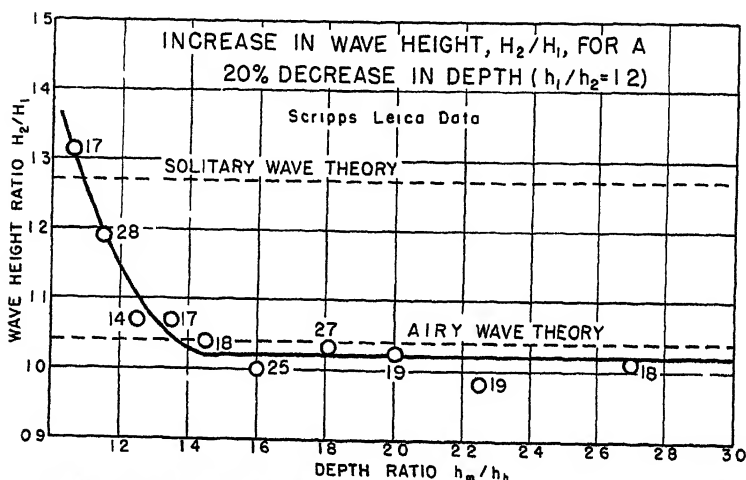


FIGURE 5. The circles denote averages based on observations along the Scripps Institution pier. The figures next to the circles indicate the number of observations which have been averaged. For small values of the depth ratio, the solitary wave theory is in better agreement with observations; for large values of the depth ratio, the Airy Wave Theory is in better agreement.

The solitary wave theory gives a much more rapid increase of wave height with decreasing depth than the Airy theory (see introduction).

In FIGURE 5, observed changes in wave height between adjacent observation stations are compared to those predicted by theory. Ratios in depth between adjacent stations are reduced to 20 per cent ($h_1/h_2 = 1.2$), and the observed changes in wave height adjusted accordingly. If, for example, h_1/h_2 equalled 1.1 for two adjacent stations, the observed difference in wave height $H_2 - H_1$ was multiplied by two, and the ratio H_2/H_1 formed accordingly. Observations are plotted against h_m/h_b where $h_m = (h_1 + h_2)/2$ is the mean depth between adjacent stations. All data pertaining to a change in depth of less than 5 per cent ($h_1/h_2 < 1.05$) have been excluded. Observations were averaged for ranges of h_m/h_b , and the number of observations upon which the average is based is shown in the figure. The heavy solid line can be taken to give the observational average. For large values of the depth ratio, the observed increase in height is consistently smaller than the cor-

responding theoretical value. This discrepancy can be ascribed to bottom friction (see TABLE 2).

For $h_m/h_b < 1.2$, the solitary wave theory is in better agreement. For $h_m/h_b > 1.2$, the Airy theory seems more applicable. Since h_m was defined as the average depth over which the transformation takes place, the solitary regime seems to extend to a depth equal to about 1.4 times the depth at the point of breaking. The corresponding value for the relative wave height is 0.35.

Wave refraction. In the case of waves traveling over an irregular bottom, an additional factor influencing wave height is associated with local convergences and divergences of wave orthogonals. By orthogonals is meant the paths described by moving points on a wave crest, traveling always normal to the axis of the crest.

The computation of the "refraction factor" is similar to the derivation leading up to EQUATION 43 with the understanding that energy is conserved over a width s of wave crest contained between adjacent orthogonals:

$$\frac{C_0}{2} \frac{E_0}{L_0} S_0 = C_b \frac{E_b}{L_b} S_b. \quad (47)$$

It follows²¹ that

$$\frac{H_b}{H_0} = \frac{1}{3.3 \sqrt[3]{H_0/L_0}} \sqrt[3]{S_0/S_b}, \quad (48)$$

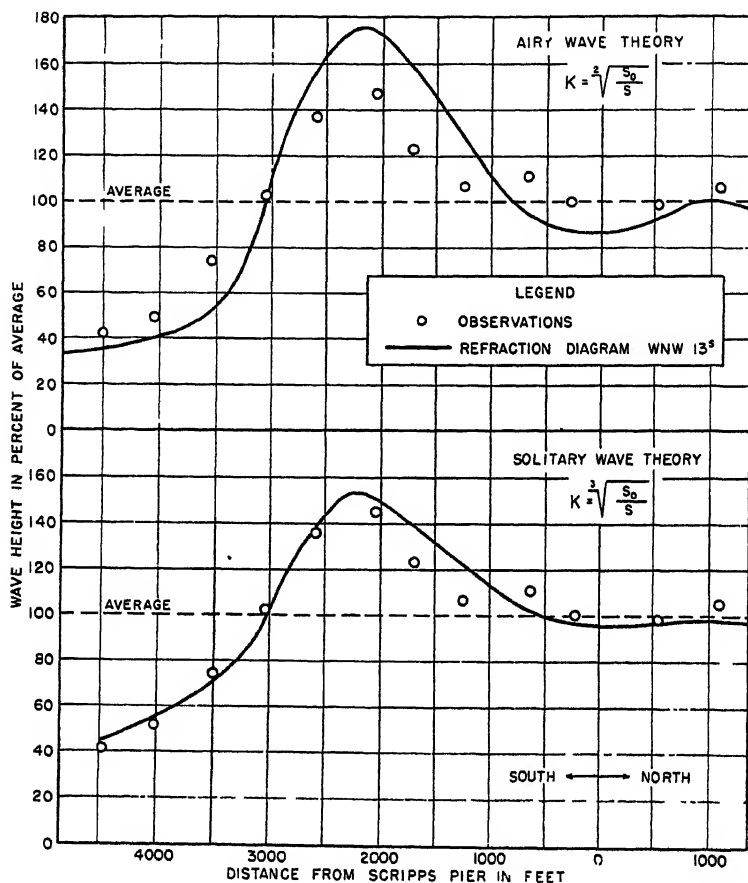
so that the breaker height along a beach must vary as the cube root of the relative spacing of orthogonals. This is borne out by observations of breaker heights along the beach adjacent to the Scripps Institution pier (FIGURE 6).

Longshore currents. Waves breaking at an angle with the coastline usually set up longshore currents inside the breaker-zone. Such currents occupy a very important role in the general study of beach erosion, velocities up to three knots having been observed.

The solitary wave theory has been applied to the study of longshore currents using two different modes of attack:²² (a) equating a certain fraction of the longshore component of the power of the breakers to the energy dissipated in the current by friction along the bottom; (b) equating the longshore component of $Q_b C_b$, the rate of momentum transfer by the breakers into the surf zone, to the frictional force exerted on the current by the ocean bottom. Both methods have led to reasonable results. FIGURE 7 shows the comparison between values computed by the momentum approach and, observations along the beach at Oceanside, California, and in the Berkeley wave tank.

The rise in the sea surface. The rise in sea surface is caused by the on-

shore momentum of the waves, similar to the manner in which longshore currents are caused by the longshore momentum of the waves. Let the mean sea surface assume a profile which gives rise to a return flow that



the beach adjacent to the solitary wave theory, gives better agreement with observations than the upper curve computed according to Airy's theory.

exactly balances the shoreward transport by the solitary waves. In the absence of frictional forces, such a return flow must be uniform from top to bottom. The elevation of the mean position of the water line on the beach above mean sea level in deep water is given closely by²⁸

$$\frac{y}{g} = \frac{2V_0^2}{g} \quad (49)$$

or, in view of (11) and (32)

$$y = \frac{13.7}{g} \left(\frac{H_b}{T} \right)^2. \quad (50)$$

The rise was measured during experiments in the Berkeley wave tank. The one set of observations for which the waves were generated in deep water, and the breaker height and depth of breaking conformed to the

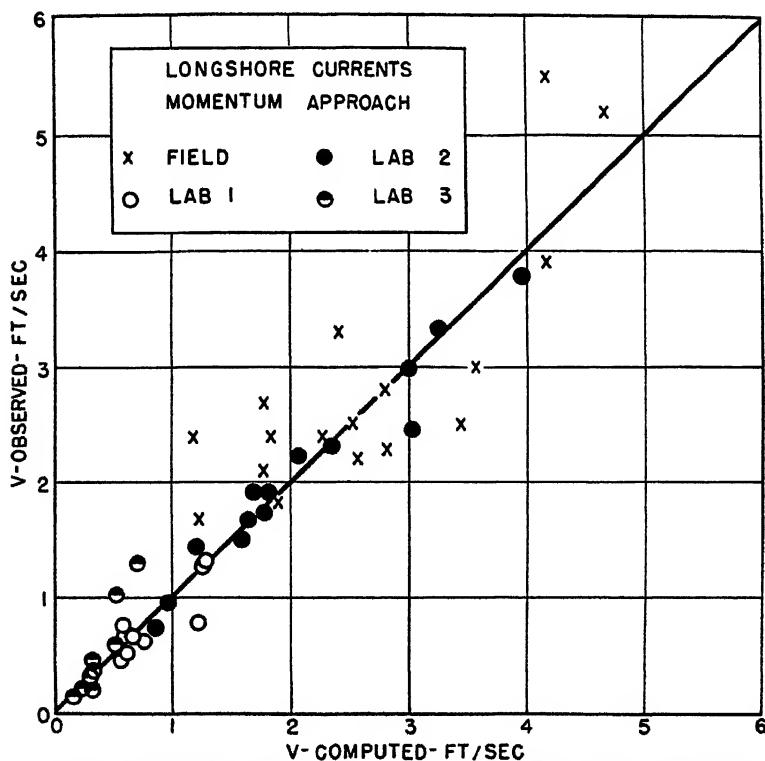


FIGURE 7. Comparisons between observed and computed velocities of longshore currents (after PUTNAM, MUNK and TRAYLOR²²).

solitary wave theory, leads to excellent agreement with theory (TABLE 4). Measurements of sea level rise along the Scripps Institution Pier are contemplated in the near future.

Discussion

The observations which have been introduced to test the solitary wave theory exhibit such a large degree of scatter that it is not possible to speak of confirmation in the sense in which physical laws can be checked in the

laboratory. The overall impression is, however, that the solitary wave theory provides a useful tool for the study of various surf phenomena.

To obtain reliable observations of some of the features here described is exceedingly difficult. In the field, extreme variability in the wave train itself is likely to obscure meaningful relationships. In the laboratory, it is difficult to avoid interference by reflected waves and surging*. Because of this very uncertainty in measurements, a theoretical framework is almost essential for establishing useful relationships. The need for some guiding theory becomes especially apparent in the important and perhaps most difficult application, the problem of sand transport. In one way or another, this problem involves all the applications which have been discussed so far.

Sand Transport: An Application of the Second Approximation

Orbital motion on a sloping beach. Sand movements must be governed largely by the orbital motion near the bottom. For the present, assume that the shoreward transport by the waves is exactly balanced by a return flow uniform from top to bottom as discussed in dealing with the rise of the

TABLE 4
MEASUREMENT OF RISE OF WATER LEVEL IN BERKELEY WAVE TANK

$H_b = 0.3$	feet
$T = 0.86$	seconds
$y_{\text{computed}} = 0.05$	feet
$y_{\text{observed}} = 0.052$	feet

sea surface. The orbital motion can then be considered composed of three parts: (a) a one-directional quasi-parabolic motion which constitutes the trajectory of particles at passage of solitary waves in still water (EQUATION 21, PLATE 4); (b) a return flow uniform from top to bottom associated with a rise in the sea surface (EQUATION 11); (c) a slight drift whose velocity and direction depends on the elevation of the particle above the bottom (EQUATION 37).

FIGURE 8 shows the three components separately and the net trajectory for a particle near the surface. The dimensions were selected to conform with an experimental determination²⁴ of the trajectory, the results of which are shown on the bottom of FIGURE 8. The trajectory was determined from motion pictures of a vertical staff which was "anchored" to the water by long vanes on its lower end. Theory and observations both reveal a net shoreward motion†, and a trajectory which has the appearance of an ellipse flattened at its bottom. The peculiar shape of the trajectory led the observer to remark²⁴ that "the true motion of the water particles was found to differ noticeably from the (Airy) theory."

* The magnitude of error can be estimated from measurements of 11 successive breakers in the Beach Erosion Board tank by means of a high speed camera.¹⁷ These varied from 0.245 feet to 0.296 feet, a total spread of 20 per cent.

† The observed drift velocity is larger than the computed drift, perhaps because of the action of the waves on the float above the vanes.

The spacing of the time marks in FIGURE 8 indicates an important feature: the shoreward motion (along the top of the trajectory) is much more rapid than the seaward motion. Numerical values, which are summarized in TABLE 5, indicate also the relative magnitudes of the wave velocity, the orbital velocities, the velocity of the return flow, and the drift velocities. The bottom drift is very slow, equalling only 1 per cent of the orbital velocity

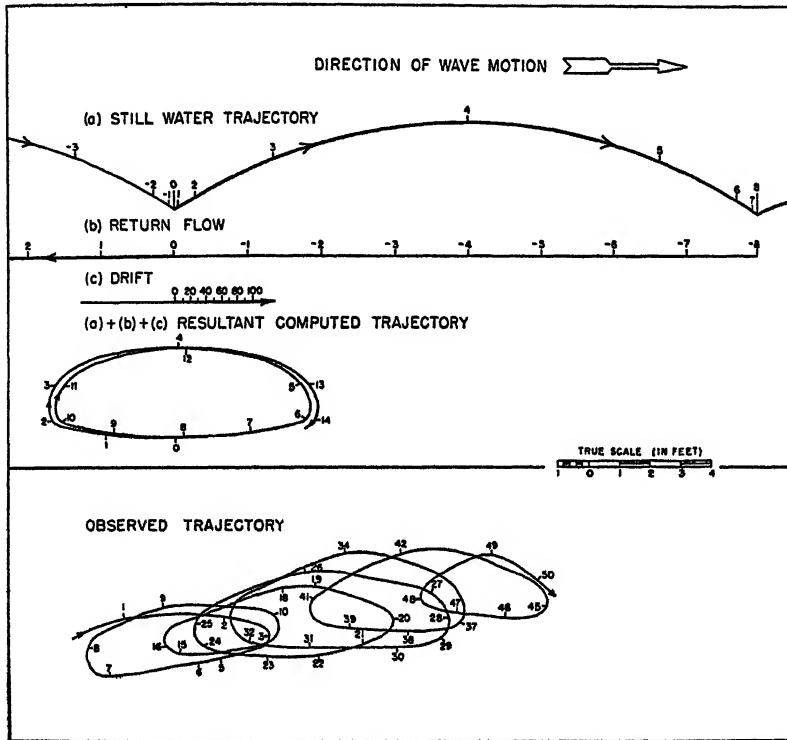


FIGURE 8. The upper section of the figure shows the three components of particle motion separately, and the trajectory which results from superposition of the three components. The lower figure shows trajectory observed during experiments conducted by the Beach Erosion Board.²⁴ The motions are drawn to scale without vertical exaggeration. Time marks (seconds) are indicated. Computed and observed trajectories both have the peculiar shape of an ellipse flattened at the bottom.

under the crest, and it cannot possibly account for the mystic "undertow" which has been reported by swimmers.

Observations described in the Beach Erosion Board report²⁴ may serve to check the order of magnitude of the bottom drift. Unfortunately the dimensions of the waves are not stated, but the assumed dimensions in TABLE 5, which were derived from the measured trajectory shown on the bottom of FIGURE 8, can be considered typical of wave conditions along the New Jersey Coast, where the measurements were conducted. The currents were determined one foot from the bottom in depth of water of 15 feet by

means of two Pegram meters, oriented at right angles to each other. Measurements were extended over 60 tidal cycles. Elimination of the tidal component gave a residual current of 0.08 feet/sec (\approx 300 feet/hour) directed off-shore! "This possibly represents a continuous slight movement of bottom-water seaward, due to an onshore surface movement caused by wave action."²⁴ The computed drift velocity of 0.05 feet/sec is of the order of magnitude of the measured residual current.

The following experiment is also reported.²⁴ "Sixty croquet balls were weighted with lead, so as to settle through salt water at about the same speed as sand particles from the adjoining beach. Thirty of these were placed in the ocean on a line parallel to and 2000 feet from the shore-line, somewhat seaward from the breaker zone. The others were dropped, during rough weather, at intervals along the Long Branch pier. *Only one of the 60, which*

TABLE 5

NUMERICAL VALUES OF WAVE CHARACTERISTICS COMPUTED ACCORDING TO SOLITARY WAVE THEORY FOR CONDITIONS SIMILAR TO THOSE UNDER WHICH MEASUREMENTS WERE CONDUCTED

Wave height H (observed)	4 feet
Water depth h (observed)	15 feet
Wave period T (observed)	8 seconds
Effective period T_{eff} (PLATE 2)	5.9 seconds
Relative wave height γ (EQUATION 3)	0.28
Horizontal displacement \bar{r} (EQUATION 17)	18.2 feet
Wave velocity C (PLATE 2)	+24.8* feet/sec
Orbital crest velocity, $u_c - \bar{v}$ (PLATE 10, EQUATION 11)	+3.9 "
Orbital trough velocity, $u_t - \bar{v}$ (PLATE 6, EQUATION 11)	-2.2 "
Return flow \bar{v} (EQUATION 11)	-2.3 "
Surface drive $v_1 - \bar{v}$ (PLATE 12)	+0.11 "
Bottom drift, $v_b - \bar{v}$ (PLATE 12)	-0.05 "

* (+) is shoreward, (-) is seaward

had lost its weight and risen to the surface, was washed ashore. It was being moved along the waterline by the waves when it was recovered. The experiment indicates that the bottom forces, at the times and places of the experiment, were either not strong enough, or not in the proper direction, to bring the balls ashore."

Shepard and Panzarini* have observed the movement, near the sea bottom, of balls carefully weighted to barely exceed the density of water. They found wide variations in the bottom drifts. In the breaker zone, three times as many observations showed seaward drift as compared to shoreward drift, but about 50 per cent were predominantly longshore. At least some of the seaward movements can be associated with rip currents. Outside the breakers, out of 30 observations about one third showed seaward components, one third shoreward components, and one third was predominantly longshore.

* Report in preparation.

In contrast to these field observations, Bagnold,¹³ in tank experiments, reports a seaward drift near the surface, a shoreward drift at the bottom.

The sand movement. In general, beaches are built up during the summer, eroded during the winter. There has, as yet, been no completely satisfactory explanation of how the sand is moved shoreward during the summer, and seaward during the winter. The first discussion of the problem which takes into account the nature of orbital motion has been given by Grant²⁵ who has proposed the following hypothesis: It has been demonstrated by theory and observations that there is a net mass transport of water in the direction of wave propagation. Experiments have furthermore shown that in the oscillatory orbital motion the "forward velocity exceeds the backward and hence in addition to net transportation there is also a *differential velocity factor* of great importance in the mechanics of onshore creep of bottom material." "If the forward motion is more rapid than the reverse movement, then some particles may be of just the proper size, weight, and shape to be moved shoreward by the high velocity in that direction but not shifted back again by the slower reverse current. On the other hand, some larger particles will not move at all while very small ones may move in both directions. In the latter case, if mass transport takes place, it would seem that the fine sediment should experience an ultimate net translation toward the shore." "The offshore movement, against the direction of wave propagation, is accomplished by the offshore mass water transport, principally that which takes place in rip currents. These currents are produced by the excess of water moved beachward by the waves."

In the previous section, it has been shown that the "differential velocity factor," here emphasized by Grant, is also consistent with the solitary wave theory. With regard to Grant's statement concerning a shoreward mass *transport*, it should be noted that this term is usually applied to the net water movement in still water. The net movement in the presence of a return flow has been termed as *drift*. The shoreward mass transport, in the foregoing sense, is much too large to be entirely balanced by outflow through rip currents. In the previous section, it was assumed that the shoreward transport is *completely* balanced by a return flow associated with a rise in sea surface. Under these conditions, the drift along the bottom is seaward, even though the mass transport is shoreward.

Whether the drift along the bottom is seaward or shoreward depends upon the extent to which this volume transport is balanced by the return flow, as compared to a compensation by rip currents (FIGURE 9). For the critical case of *zero bottom drift*, the shoreward transport of volume equals $h^2\bar{R}/T$, the return flow h^2R_0/T , and the mean outward movement per unit length of beach due to rip currents

$$\frac{h^2(\bar{R} - R_0)}{T}. \quad (51)$$

If the mean outflow due to rip currents exceeds the amount given in EQUATION (51), then the bottom drift is shoreward; if it is less, the bottom drift is seaward. Should the latter condition prevail, the following explanation for the sand movement might be suggested: During the summer, when waves are generally long and low, the "differential velocity factor" is particularly pronounced, and the sand particles are moved only by the relatively rapid shoreward component of orbital motion. In this manner, sand par-

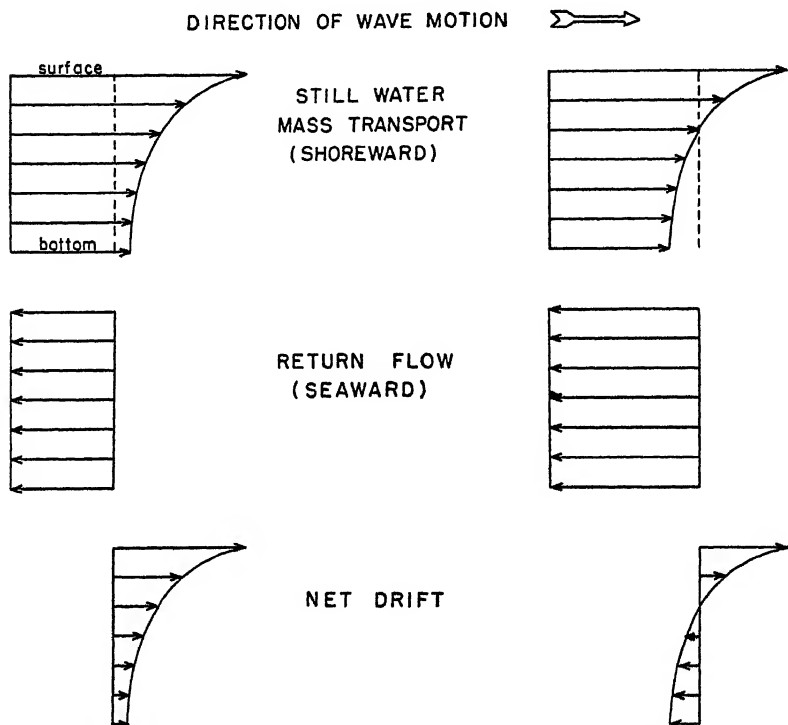


FIGURE 9. The figures to the right correspond to the case where the mass transport shoreward is entirely compensated for by a uniform return flow. In this case, the drift along the bottom is seaward. The figure to the left corresponds to conditions under which a substantial portion of the shoreward mass transport is returned in rip currents. The net drift along the bottom is shoreward.

ticles move shoreward even though the water along the bottom drifts seaward. In the winter, waves are generally high and short, turbulence large, and the differential velocity factor small. The fine sand particles remain in suspension, the larger particles roll back and forth to an almost equal extent. In either case, the sand gradually moves seaward with the mean drift of water along the bottom.

The foregoing remarks are intended as a suggestion of a mechanism for sand transport which is somewhat different from the one proposed by Grant. A great deal of work remains to be done before any definite conclusions can

be reached. To the extent to which the problem of sand transport involves the seaward drift of water along the bottom it depends on a higher degree of approximation of the solitary wave theory than do the previous applications.

Bibliography

1. U. S. NAVY HYDROGRAPHIC OFFICE. 1944. Wind waves and swell: Principles in forecasting. H. O. Misc. 11: 275.
2. SVERDRUP, H. U., & W. H. MUNK. 1946. Empirical and theoretical relations between wind, sea and swell. Trans. Am. Geophys. Union. 27: 823-827.
3. BURNSIDE, W. 1915. On the modification of a train of waves as it advances into shallow water. Proc. Lond. Math. Soc. II. 14: 131.
4. SVERDRUP, H. U., & W. H. MUNK. 1946. Theoretical and empirical relations in forecasting breakers and surf. Trans. Am. Geophys. Union. 27: 828-836.
5. BEACH EROSION BOARD. 1942. Technical Report No. 2. A summary of the theory of oscillatory waves. United States Government Printing Office.
6. U. S. NAVY HYDROGRAPHIC OFFICE. 1944. Breakers and surf, principles in forecasting. H. O. 234.
7. MUNK, W. H. 1945. Supplement to breakers and surf, principles in forecasting. S. I. O. Wave Report No. 49 (unpublished).
8. KELLER, J. B. The solitary wave and periodic waves in shallow water. N. Y. Acad. Sci. 51: (3) 345.
9. KEULEGAN, A. H. & G. W. PATTERSON. Mathematical theory of irrotational translation waves. J. Res. Nat. Bur. Standards. 24: 48-101.
10. LAMB, H. 1932. Hydrodynamics. Cambridge Univ. Press (6th Edition) London. 258.
11. McCOWAN, J. 1891. On the solitary wave. Phil. Mag. 32: (5) 45-58.
12. McCOWAN, J. 1894. On the highest wave of permanent type. Phil. Mag. 38: (5) 351-357.
13. BAGNOLD, R. A. 1947. Sand movement by waves. J. Inst. Civ. Eng. 27: 447-469.
14. SCRIPPS INSTITUTION OF OCEANOGRAPHY. 1944. Waves in shallow water (I). Change in wave height along Scripps Institution pier with special emphasis on breaker characteristics. Wave Report No. 13 (unpublished).
15. SCRIPPS INSTITUTION OF OCEANOGRAPHY. 1944. Effect of bottom slope on breaker characteristics as observed along the Scripps Institution pier. Wave Report No. 24 (unpublished).
16. EDMONDSON, T. E., & G. L. CLARKE. 1945. Measurements of waves, surf and shore conditions on Martha's Vineyard and Cape Cod. Woods Hole Oceanographic Institution (unpublished).
17. REYNOLDS, K. C. 1945. Report on the laboratory investigation of characteristics of waves approaching beaches. Woods Hole Oceanographic Institution (unpublished).
18. UNIVERSITY OF CALIFORNIA. 1945. Preliminary report on model studies on the transformation of waves in shallow water. Report No. HE-116-106 (unpublished).
19. GAILLARD, D. D. 1935. Wave action in relation to engineering structures. The Engineer School, Fort Belvoir, Virginia.
20. MUNK, W. H., & M. A. TRAYLOR. 1945. Height of breakers and depth of breaking. S. I. O. Wave Report No. 47 (unpublished).
21. MUNK, W. H., & M. A. TRAYLOR. 1947. Refraction of ocean waves: a process linking underwater topography to beach erosion. J. Geol. 15: 1-26.
22. PUTNAM, J. A., W. H. MUNK, & M. A. TRAYLOR. The prediction of longshore currents. Trans. Am. Geophys. Union. In press.
23. MUNK, W. H., & P. OSBORN. (In preparation).
24. BEACH EROSION BOARD. 1933. Office of the Chief of Engineers, U. S. Army. Interim Report. (15 April).
25. GRANT, U. S. 1943. Waves as a sand transporting agent. Am. J. Science, 241: 117-123.

Appendix 1: Computation of Drift

General Equations. Let the parameters m and a be defined by

$$mh = M, \quad ma = N$$

where M and N are non-dimensional numbers given by EQUATIONS 27 and 28. EQUATIONS 25 and 26, for the horizontal and vertical components of orbital velocity, then become

$$u = \frac{d\xi}{dt} = Cma \frac{1 + \cos mz' \cosh mx'}{(\cos mz' + \cosh mx')^2}, \quad (52)$$

$$w = \frac{d\zeta}{dt} = Cma \frac{\sin mz' \sinh mx'}{(\cos mz' + \cosh mx')^2}, \quad (53)$$

where

$$x' = x + \xi - Ct, \quad z' = z + \zeta \quad (54)$$

are the instantaneous coordinates* of motion of a particle initially at x, z . The difficulty is that (52) and (53) cannot be integrated independently of one another, since, for example, $u = f(x', z')$ and z' depends upon the vertical motion.

McCowan shows that by a special device the vertical component can be integrated directly, and finds

$$\zeta = a \frac{\sin mz'}{\cos mz' + \cosh mx'}. \quad (55)$$

To obtain ξ it is necessary to proceed by successive approximations. In general

$$\frac{d\xi}{dt} = \frac{\partial \xi}{\partial x'} \frac{\partial x'}{\partial t} + \frac{\partial \xi}{\partial z'} \frac{\partial z'}{\partial t},$$

or, in view of (54)

$$\frac{d\xi}{dt} = \frac{\partial \xi}{\partial x'} (u - C) + \frac{\partial \xi}{\partial z'} w. \quad (56)$$

Since the motion is principally directed along the x -axis, the second term on the right side of (56) is small. Furthermore, $u \ll C$, and (56) becomes approximately

$$\frac{d\xi}{dt} \doteq -C \frac{d\xi}{dx'}, \quad (57)$$

or

$$\xi = -\frac{1}{C} \int \frac{d\xi}{dx'} dx'.$$

* In the main body of the paper the initial coordinates are designated by x_0, z_0 , the displacements from their initial position by $r' + \bar{r}/2$ and z' .

The integration is best performed by replacing x' by θ , where, by definition

$$\begin{aligned}\sin m\theta &= \frac{w}{q} = \frac{\sin mz' \sinh mx'}{\cos mz' + \cosh mx'}, \\ \cos m\theta &= \frac{u}{q} = \frac{1 + \cos mz' \cosh mx'}{\cos mz' + \cosh mx'},\end{aligned}\quad (58)$$

and $q^2 = u^2 + w^2$. Differentiating either expression gives

$$\begin{aligned}dx' &= \frac{\cos mz' + \cosh mx'}{\sin mz'} d\theta \\ &= \frac{\sin mz'}{\cos m\theta - \cos mz'} d\theta.\end{aligned}\quad (59)$$

Substituting (52) and (59) into (57) gives, after some substitution,

$$\xi = -\frac{1}{C} \int \frac{Cma}{\sin mz'} \cos m\theta d\theta = -\frac{a \sin \theta}{\sin mz'} + \text{const} \quad (60)$$

and reverting to original coordinates

$$\xi = -a \frac{\sinh mx'}{\cos mz' + \cosh mx'} + \text{const.} \quad (61)$$

To correct for the approximation in writing (57), a correction term τ is added to EQUATION 61 which can be evaluated by referring to the exact EQUATION 56:

$$\xi = -a \frac{\sinh mx'}{\cos mz' + \cosh mx'} + \tau + \text{const.} \quad (62)$$

Differentiating and substituting into (56) gives, after some simplification,

$$\frac{d\tau}{dt} = \frac{Cm^2 a^2}{(\cos mz' + \cosh mx')^2}, \quad (63)$$

or, in terms of θ ,

$$\frac{d\tau}{dt} = \frac{Cm^2 a^2}{\sin mz'} (\cos m\theta - \cos mz')^2. \quad (64)$$

The approximation process is now repeated. Proceeding as in (57) by "partial integration"

$$\tau = -\frac{1}{C} \int \frac{\partial \tau}{\partial t} dx' = -\frac{ma^2}{\sin^3 mz'} (\sin m\theta - m\theta \cos mz') + \text{const}, \quad (65)$$

and reverting to original coordinates

$$\begin{aligned}\tau &= -\frac{ma^2}{\sin^3 mz'} \left\{ \frac{\sin mz' \sinh mx'}{\cos mz' + \cosh mx'} \right. \\ &\quad \left. - \sin^{-1} \left(\frac{\sin mz' \sinh mx'}{\cos mz' + \cosh mx'} \right) \cos mz' \right\} + \text{const.}\end{aligned}\quad (66)$$

Adding τ_1 to (66) as a second correction factor, differentiating according to (56) and substituting into (63) gives, after much manipulation,

$$\frac{d\tau_1}{dt} = \frac{Cm^3 a^3}{\sin^5 mz'} (\cos m\theta - \cos mz') \{ \cos m\theta (\cos m\theta - \cos mz')^2 + m\theta \sin m\theta (1 + 2 \cos^3 mz') + \sin^2 m\theta (\cos m\theta - 4 \cos mz') \} \quad (67)$$

and by partial integration

$$\begin{aligned} \tau_1 &= -\frac{1}{C} \int \frac{\partial \tau_1}{\partial t} dx' \\ &= -\frac{m^2 a^3}{\sin^5 mz'} \left\{ \frac{\sin^3 m\theta}{3} + \sin m\theta \left(\frac{5}{3} + 3 \cos^2 mz' \right) - 3m\theta \cos mz' \right. \\ &\quad \left. + \frac{\sin m\theta \cos m\theta}{3} (\cos m\theta + 3 \cos mz') - m\theta \cos m\theta (1 + 2 \cos^2 mz') \right\} \\ &\quad + \text{const.} \end{aligned} \quad (68)$$

To find the total horizontal displacement r due to a wave, EQUATIONS 60, 65, and 68 are integrated from $t = -\infty$ to $t = +\infty$; the constants then drop out, and the other terms have the following values at the respective limits of integration:

$$\begin{aligned} t &= -\infty & t &= +\infty \\ x' &= +\infty & x' &= -\infty \\ \sinh mx' &= +\infty & \sinh mx' &= -\infty \\ \cosh mx' &= +\infty & \cosh mx' &= +\infty \\ \tanh mx' &= +1 & \tanh mx' &= -1 \\ \sin m\theta &= +\sin mz & \sin m\theta &= -\sin mz \\ \cos m\theta &= +\cos mz & \cos m\theta &= +\cos mz \\ m\theta &= mz & m\theta &= -mz \end{aligned} \quad (54)$$

Then

$$\begin{aligned} r &= \int_{-\infty}^{+\infty} \xi dt \\ &= 2a \left\{ 1 + ma \left[\frac{\sin mz - mz \cos mz}{\sin^3 mz} \right] \right. \\ &\quad \left. + 2m^2 a^2 \left[\frac{(3 - \sin^2 mz)(\sin mz - mz \cos mz) + \sin^3 mz}{\sin^5 mz} \right] \right\}. \end{aligned} \quad (69)$$

Reverting to non-dimensional coordinates

$$R = 2 \frac{N}{M} \left\{ 1 + N \left[\frac{\sin MZ - MZ \cos MZ}{\sin^3 MZ} \right] + 2N^2 \left[\frac{(3 - \sin^2 MZ)(\sin MZ - MZ \cos MZ) + \sin^3 MZ}{\sin^5 MZ} \right] \right\}. \quad (70)$$

For small values of z EQUATION 70 can be expanded in a Taylor series. Let $f_1(MZ)$ and $f_2(MZ)$ denote respectively the terms in the brackets of EQUATION 70. Then, after very many steps,

$$\begin{aligned} f_1(0) &= 1/3 & f_1'(0) &= 0 & f_1''(0) &= 4/15 & f_1'''(0) &= 0 \\ f_2(0) &= 1/15 & f_2'(0) &= 0 & f_2''(0) &= 4/21 & f_2'''(0) &= 0 \end{aligned}$$

and

$$R = 2 \frac{N}{M} \{ 1 + N[\frac{1}{3} + \frac{2}{15}(MZ)^2 + \dots] + 2N^2[\frac{1}{15} + \frac{2}{31}(MZ)^2 + \dots] + \dots \}. \quad (71)$$

Displacement along the bottom. Along the bottom $Z = 0$, and (71) becomes

$$R_0 = 2 \frac{N}{M} \{ 1 + \frac{1}{3}N + \frac{2}{15}N^2 + \dots \}. \quad (72)$$

But for this special case the motion is confined to the horizontal, and R_0 can be found directly without approximation. Substituting from EQUATIONS 14 and 31 gives

$$\begin{aligned} R_0 &= 2N \int_0^\infty \frac{1}{1 + N + \cosh MX} dX \\ &= 2 \frac{N}{M} \beta \operatorname{cosec} \beta, \end{aligned} \quad (73)$$

where $\cos \beta = 1 - N$. Expanding by Taylor's theorem leads to a series of which the first three terms are in agreement with (72).

The mean displacement. The mean displacement is defined by the equation

$$\bar{R} = \int_0^1 R dZ$$

where $R(Z)$ is given by EQUATION 70. Since the degree of approximation of (70) is not known, it is preferable to determine \bar{R} from

$$\bar{R} = \frac{Q}{h^2} \quad (19)$$

where

$$Q = 2h^2 \int_0^\infty \mu \, dX \quad (9)$$

is the volume per unit crest width. McCowan gives

$$\mu = \frac{N}{M} \frac{\sin M(1 + \mu)}{\cos M(1 + \mu) + \cosh MX}, \quad (29)$$

which cannot be integrated directly because μ appears on both sides of the equation. Following McCowan's suggestion of expanding (29) by Lagrange's theorem and integrating term by term, one obtains, after many steps, the series

$$\bar{R} = 2 \frac{N}{M} (1 + \frac{1}{3}N + \frac{2}{15}N^2 + \frac{2}{35}N^3 + \frac{2}{45}M^2N + \frac{19}{594}M^2N^2 + \dots) \quad (74)$$

of which the first two terms were already given by McCowan. Unfortunately the series is found to converge slowly, and higher order terms become exceedingly complex.

The method adapted here makes use of the close approximation of Boussinesq's equation

$$\mu = \gamma \operatorname{sech}^2 \left(\sqrt{\frac{3\gamma}{4}} X \right) \quad (7)$$

to the more accurate EQUATION 29. Letting the subscripts "B" and "M" refer to Boussinesq's and McCowan's solutions, respectively, one can write

$$Q_M = Q_B + 2h^2 \int_0^\infty (\mu_M - \mu_B) \, dX \quad (75)$$

where μ_M and μ_B are given by (29) and (7) respectively, and where

$$Q_B = 4h^2 \sqrt{\frac{\gamma}{3}}. \quad (9)$$

The integration was performed graphically for eight values of γ . Since $Q_M - Q_B$ is small, this method permits the determination of Q_M and \bar{R} to four significant figures.

Final Computations of $R(Z)$. Let

$$R_1 = 2 \frac{N}{M}$$

$$R_2 = 2 \frac{N}{M} \{1 + N[f_1(MZ)]\}$$

$$R_3 = 2 \frac{N}{M} \{1 + N[f_1(MZ)] + 2N^2[f_2(MZ)]\}$$

represent successive degrees of approximation to EQUATION 70. The three approximations are plotted in FIGURE 10 for the special case $\gamma = 0.5$. Since R_3 still differs appreciably from R_2 , further refinements are desirable, but the rapidly growing complexity of the expansion makes this very difficult.

FIGURE 10 also shows the value of \bar{R} obtained by the graphical method discussed above, and the exact value of the bottom displacement, R_0 , as found from EQUATION 73. In drawing the final curves $R(Z)$ we are guided by the following three considerations: (a) the curves must go through R_0 ; (b) the average value of $R(Z)$ must equal \bar{R} ; (c) the shape of the curves must resemble R_3 , and, to a lesser extent R_2 .

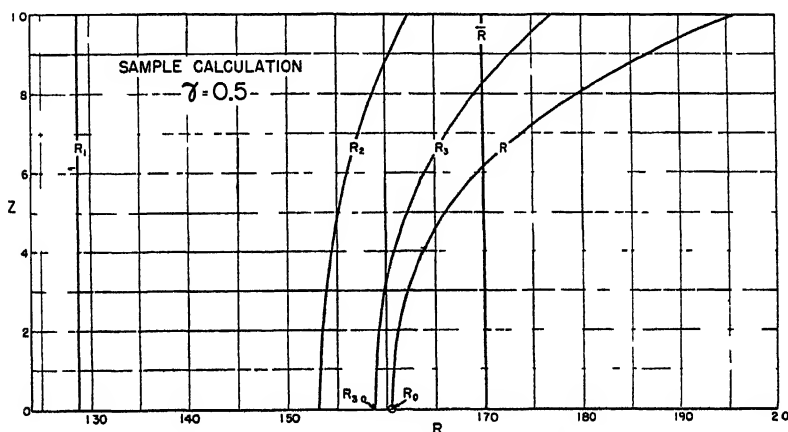


FIGURE 10. Sample calculation of net horizontal particle displacements. The lines marked R_1 , R_2 , R_3 and R represent successive stages of approximation. \bar{R} represents the mean horizontal displacement and R_0 the horizontal displacement along the bottom of the last approximation.

It is clear from FIGURE 10 that the curve $R_3(Z)$ does not satisfy (a) and (b), and that the final curve $R(Z)$ must lie considerably to the right of R_3 . The difference $R_3 - R_2$ obeys approximately the cubic equation

$$R_3 - R_2 = a' + b'Z^3.$$

Assume that, in a similar manner,

$$R = R_3 + a + bZ^3. \quad (76)$$

Condition (a) requires that

$$a = R_0 - R_{3,0}, \quad (77)$$

where $R_{3,0}$ is the value of R_3 at the bottom. Condition (b) requires that

$$\int_0^1 (R - R_3) dZ = 0,$$

or, in view of (76)

$$b = -4(a + \Gamma) \quad (78)$$

where

$$\Gamma = \int_0^1 (R_s - \bar{R}) dZ$$

can be determined by graphical integration. In this manner, the constants a and b were determined, and $R(Z)$ computed for various values of γ (PLATE 12).

For the special case of breaking waves ($\gamma = 0.78$), the procedure followed was similar to the one described, except that the profile near the breaking crest was computed according to (33) instead of (29).

Appendix 2

m	T	H_0	H_b	h_b	$\frac{H_0}{L_0}$	$\frac{H_b}{H_0}$	$\frac{h_b}{H_b}$	$\frac{s}{L_0}$	$\left(\frac{H_b/H_0}{h_b/H_b}\right)$
Berkeley Wave Tank									
1:13.9	.087	.351	.324	.412	.092	.92	1.27		1.01
1:13.9	1.15	.302	.323	.351	.045	1.07	1.09		1.07
1:13.9	1.22	.284	.326	.312	.037	1.15	0.96		1.07
1:13.9	1.50	.210	.310	.270	.018	1.46	0.89		1.15
1:13.9	1.54	.199	.286	.273	.016	1.44	0.95		1.11
1:13.9	1.97	.147	.270	.234	.007	1.83	0.87		1.15
1:18.5	.86	.335	.301	.454	.088	.89	1.46		.97
1:18.5	.965	.318	.299	.363	.067	.94	1.21		1.01
1:18.5	1.34	.240	.271	.245	.026	1.13	0.90		.98
1:18.5	1.50	.205	.260	.245	.018	1.27	0.94		1.01
1:18.5	1.97	.134	.224	.207	.007	1.67	0.92		1.04
1:11.1	1.05	.335	.327	.470	.059	0.98	1.44		1.04
1:11.1	1.09	.315	.320	.469	.051	1.03	1.47		1.05
1:11.1	1.35	.270	.323	.476	.029	1.23	1.48		1.10
1:11.1	1.50	.222	.308	.474	.019	1.39	1.46		1.11
1:11.1	1.98	.150	.286	.387	.007	1.91	1.35		1.19
Scripps Leica Data, Type I									
13.7		7.4	7.4	.0042	1.85	1.00	1.10		.98
12.0		4.8	5.3	.0042	1.55	1.10	1.17		.83
13.3		5.4	7.6	.0045	1.35	1.41	1.15		.74
12.7		7.4	9.1	.0051	1.76	1.23	.97		1.00
12.2		6.4	7.2	.0051	1.64	1.13	1.02		.93
10.2		4.0	5.1	.0051	1.48	1.28	1.05		.84
11.6		7.0	8.7	.0052	1.95	1.24	.95		1.11
12.0		5.8	7.5	.0053	1.49	1.29	1.10		.85
11.5		7.6	9.1	.0066	1.46	1.20	.98		.90
10.0		5.6	6.6	.0082	1.34	1.18	.93		.88
10.0		6.6	6.6	.0082	1.57	1.00	.93		1.03
10.0		5.6	6.5	.0084	1.19	1.16	.96		.78

Appendix 2—Continued

m	T	H_0	H_b	h_b	$\frac{H_0}{L_0}$	$\frac{H_b}{H_0}$	$\frac{h_b}{H_b}$	$\frac{s}{L_0}$	$\left(\frac{H_b/H_0}{H_b/H_0}\right)$
Scripps Leica Data, Type I—Continued									
	11.2		6.6	7.6	.0084	1.22	1.15	.95	.80
	9.2		4.6	5.7	.0084	1.28	1.24	.92	.84
	9.0		5.2	7.3	.0085	1.49	1.40	.78	.98
	10.2		6.6	9.8	.0087	1.44	1.49	.97	.96
	10.5		8.6	10.3	.0090	1.69	1.20	.79	1.13
	10.0		6.6	7.2	.0094	1.37	1.09	.88	.92
	9.5		4.8	7.1	.0094	1.12	1.48	1.08	.76
	9.6		7.0	7.3	.0098	1.71	1.04	1.06	1.17
	9.5		5.8	6.6	.0102	1.23	1.14	.88	.85
	9.4		6.4	8.2	.0107	1.33	1.26	.79	.93
	9.5		8.0	9.8	.0109	1.60	1.23	.82	1.12
	9.6		7.2	9.4	.0109	1.41	1.31	1.00	.99
	10.3		8.0	9.8	.0113	1.31	1.22	.80	.93
	10.5		9.0	12.7	.0114	1.41	1.41	.68	1.00
	10.5		7.2	10.8	.0116	1.22	1.50	.91	.87
	9.6		8.8	12.2	.0119	1.57	1.39	.60	1.12
	9.8		7.0	8.8	.0122	1.17	1.26	.80	.84
	8.1		4.6	6.0	.0124	1.12	1.30	.80	.81
	10.3		9.4	10.0	.0124	1.40	1.06	.73	1.01
	9.0		6.8	9.7	.0132	1.26	1.43	.80	.93
	9.4		8.4	8.8	.0138	1.36	1.05	.54	1.01
	9.0		6.8	6.0	.0139	1.19	.88	.89	.89
	7.7		5.0	7.9	.0140	1.19	1.58	.78	.89
	9.0		9.8	11.0	.0141	1.66	1.12	.59	1.24
	8.5		6.0	7.9	.0141	1.15	1.32	.72	.86
	8.8		7.8	8.1	.0149	1.35	1.04	.73	1.02
	8.8		7.2	8.4	.0151	1.22	1.17	.72	.93
	10.0		8.0	12.1	.0151	1.29	1.51	.68	.98
	8.0		5.8	6.4	.0153	1.18	1.10	.85	.90
	7.2		6.6	8.4	.0158	1.61	1.27	.44	1.24
	9.0		8.0	8.9	.0171	1.13	1.11	.72	.89
	8.0		6.2	7.0	.0175	1.12	1.13	.68	.88
	9.2		8.6	12.2	.0176	1.13	1.42	.54	.89
	8.8		7.8	9.2	.0179	1.11	1.18	.67	.89
	8.5		9.8	9.1	.0181	1.44	.93	.82	1.14
	8.0		7.2	8.6	.0194	1.14	1.20	.59	.92
	7.5		7.8	10.5	.0197	1.37	1.35	.35	1.11
	9.0		11.4	12.1	.0200	1.39	1.06	.55	1.13
	8.2		10.0	11.3	.0235	1.25	1.13	.42	1.07
	7.5		9.0	11.2	.0239	1.34	1.24		1.14
	7.2		8.0	9.9	.0239	1.29	1.24	.30	1.10
	8.0		9.4	9.9	.0243	1.18	1.05	.50	1.02
	6.5		7.2	8.2	.0284	1.01	1.14	.32	.90
	7.8		11.0	14.6	.0316	1.12	1.33	.07	1.04

Appendix 2—Continued

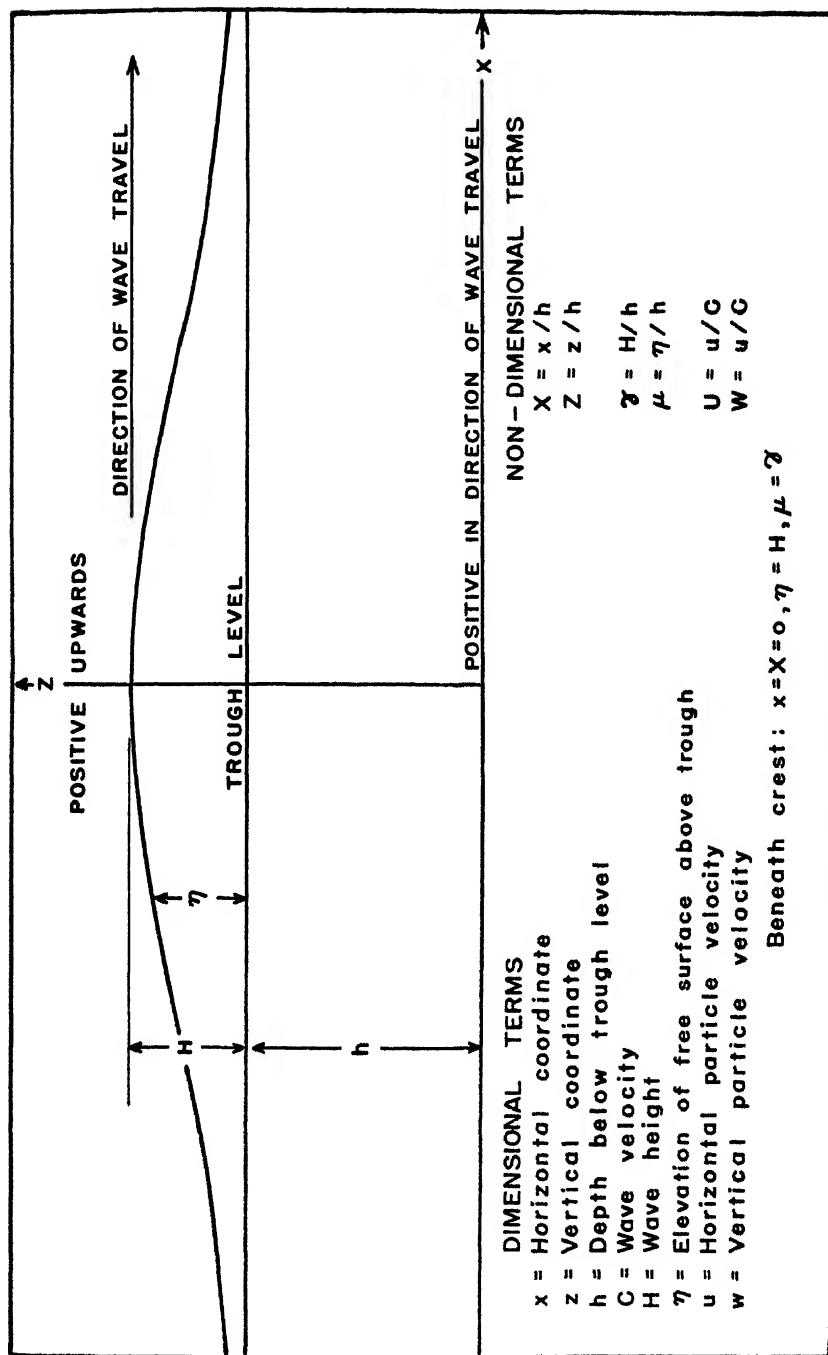
m	T	Π_0	H_b	h_b	$\frac{H_0}{L_0}$	$\frac{H_b}{H_0}$	$\frac{h_b}{H_b}$	$\frac{s}{L_0}$	$\left(\frac{H_b/H_0}{H_b/H_0}\right)$
Scripps Leica Data Type II									
	13.0		5.8	10.3	.0046	1.45	1.77	1.10	.80
	12.5		6.4	9.9	.0061	1.31	1.55	1.22	.79
	12.0		7.6	10.6	.0074	1.38	1.39	1.05	.87
	10.5		7.2	12.2	.0088	1.47	1.69	.95	.98
	11.2		8.2	11.4	.0091	1.39	1.39	1.16	.93
	10.0		6.0	11.6	.0092	1.28	1.93	1.06	.86
	10.0		6.0	10.9	.0093	1.25	1.82	.91	.84
	8.8		4.2	5.5	.0098	1.11	1.31	1.12	.76
	9.3		7.8	11.9	.0100	1.77	1.53	.69	1.21
	9.6		8.4	11.3	.0106	1.68	1.35	1.08	1.16
	10.5		7.6	11.5	.0107	1.27	1.51	1.08	.89
	10.0		6.9	12.2	.0110	1.24	1.77	.99	.87
	9.5		7.8	9.3	.0118	1.41	1.19	1.00	1.01
	8.9		8.2	11.8	.0125	1.64	1.44	1.08	1.19
	9.0		7.6	11.0	.0129	1.43	1.40	.90	1.04
	9.0		7.4	10.4	.0137	1.32	1.40	.89	.99
	8.0		7.6	11.3	.0191	1.25	1.49	.58	1.01
	7.0		9.0	10.9	.0260	1.38	1.21	.52	1.20
Beach Erosion Board Laboratory Tank Data									
1:6.3	.97	.107	.116	.159	.0230	1.08	1.37	.41	.92
1:6.3	1.08	.165	.170	.191	.0284	1.03	1.12	.42	.92
1:6.3	1.08	.123	.141	.143	.0207	1.15	1.02	.48	.95
1:20.4	1.08	.163	.213	.181	.0271	1.31	.85	.41	1.16
1:20.4	1.08	.125	.168	.142	.0209	1.35	.84	.54	1.13
1:20.4	.96	.117	.141	.147	.0249	1.21	1.04	.38	1.05
1:33.3	1.03	.119	.140	.200	.0218	1.18	1.43	.45	.99
1:33.3	1.03	.160	.178	.260	.0296	1.11	1.46	1.72	1.01
1:33.3	.85	.100	.107	.150	.0273	1.07	1.40	1.96	.95
1:33.3	1.03	.087	.113	.153	.0159	1.13	1.35	.73	.88
1:6.3	1.08	.235	.211	.334	.0394	.90	1.58	.28	.88
1:6.3	1.08	.192	.168	.270	.0331	.88	1.61	.34	.82
1:20.4	1.08	.187	.218	.265	.0315	1.17	1.22	.11	1.08
1:20.4	.97	.162	.204	.220	.0352	1.26	1.08	.05	1.20
1:33.3	1.03	.180	.166	.267	.0331	.92	1.61	1.69	.86
1:33.3	.85	.130	.123	.171	.0354	.95	1.39	1.78	.90
1:33.3	.75	.100	.108	.168	.0350	1.08	1.55	1.37	1.03
1:6.3	.97	.193	.179	.270	.0413	.93	1.51	.27	.92
1:6.3	.75	.113	.110	.143	.0402	.97	1.30	.31	.95
1:6.3	.74	.127	.129	.175	.0446	1.02	1.35	.24	1.03
1:20.4	1.08	.240	.275	.343	.0400	1.15	1.25	.87	1.13
1:20.4	.95	.214	.226	.290	.0453	1.06	1.28	.79	1.07
1:20.4	.73	.120	.143	.176	.0422	1.19	1.23	.78	1.19
1:33.3	.85	.145	.133	.207	.0403	.92	1.56	1.42	.90

Appendix 2—Continued

m	T	H_0	H_b	h_b	$\frac{H_0}{L_0}$	$\frac{H_b}{H_0}$	$\frac{h_b}{H_b}$	$\frac{s}{L_0}$	$\left(\frac{H_b/H_0}{H_b/H_0}\right)$
Beach Erosion Board Laboratory Tank Data—Continued									
1:33.3	.75	.140	.101	.141	.0496	.72	1.40	1.67	.75
1:6.3	.96	.250	.262	.334	.0529	1.05	1.27	.18	1.08
1:6.3	.74	.170	.159	.191	.0592	.94	1.20	.21	.99
1:6.3	1.09	.330	.307	.365	.0546	.93	1.19	.25	.97
1:20.4	1.08	.324	.329	.456	.0540	1.02	1.38	.49	1.07
1:20.4	.97	.240	.259	.320	.0500	1.08	1.24	.68	1.11
1:20.4	.75	.166	.145	.181	.0566	.87	1.25	.78	.91
1:20.4	.74	.165	.156	.206	.0554	.95	1.32	.63	.99
1:20.4	.75	.170	.159	.211	.0576	.94	1.33	.58	.98
1:6.3	.97	.299	.311	.413	.0626	1.04	1.33	.085	1.09
1:6.3	1.09	.397	.398	.556	.0651	1.00	1.40	.055	1.06
1:20.4	1.08	.402	.428	.612	.0670	1.06	1.43	.0	1.13
1:20.4	.97	.296	.296	.370	.0652	.95	1.25	.47	1.01
Field Data Collected by Woods Hole Oceanographic Institution									
					.0055	1.60	.78		.92
					.0072	1.37	1.24		.86
					.0074	1.44	1.76		.91
					.0082	1.66			1.08
					.0083	1.80			1.17
					.0083	1.45			.95
					.0104	1.31	1.05		.91
					.0110	1.77	.84		1.25
					.0114	1.36	1.35		.96
					.0114	1.28	.86		.91
					.0115	1.68	1.09		1.20
					.0116	1.55	1.02		1.11
					.0120	1.90			1.37
					.0120	1.75			1.26
					.0125	1.35			.98
					.0126	1.00	.98		.73
					.0130	1.49			1.10
					.0130	1.29			.95
					.0132	1.36			1.00
					.0135	1.18	1.38		.87
					.0139	1.70	.93		1.27
					.0144	1.30	1.29		.98
					.0161	1.47			1.14
					.0161	1.17			.91
					.0170	1.00			.78
					.0170	.97			.76
					.0176	1.06	1.11		.84
					.0188	1.47			1.18
					.0188	1.37			1.10
					.0188	1.33			1.07

Appendix 2—Concluded

m	T	H_o	H_b	h_b	$\frac{H_o}{L_o}$	$\frac{H_b}{H_o}$	$\frac{h_b}{H_b}$	$\frac{a}{L_o}$	$\left(\frac{H_b H_o}{H_b H_o}\right)$
Field Data Collected by Woods Hole Oceanographic Institution— <i>Continued</i>									
					.0188	1.28			1.03
					.0188	1.25			1.01
					.0204	1.58	1.06		1.30
					.0210	1.55			1.28
					.0211	1.38			1.14
					.0214	1.17			.98
					.0235	1.50	.85		1.27
					.0284	.97	1.00		.87

PLATE 1. Notation. ($\eta, \delta, -W = u/C$ should read $W = w/C$)

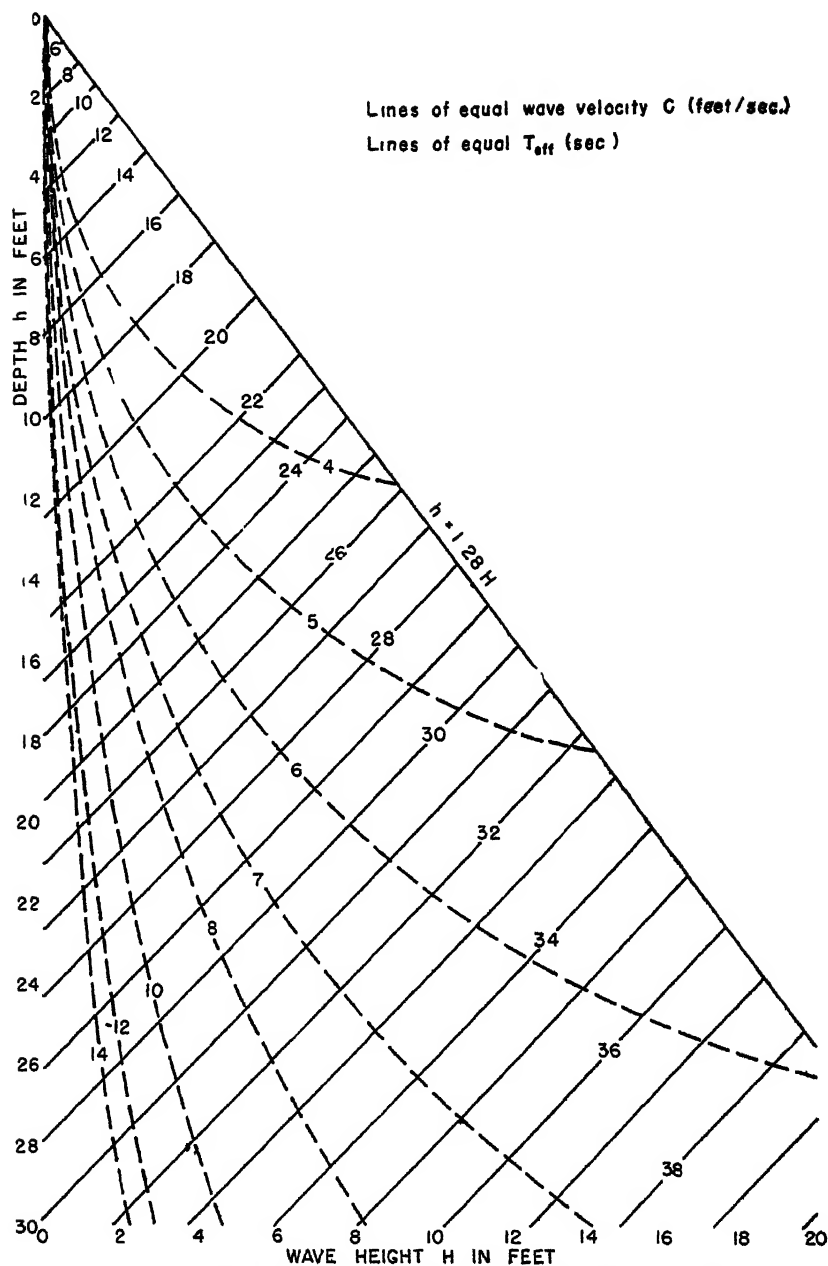
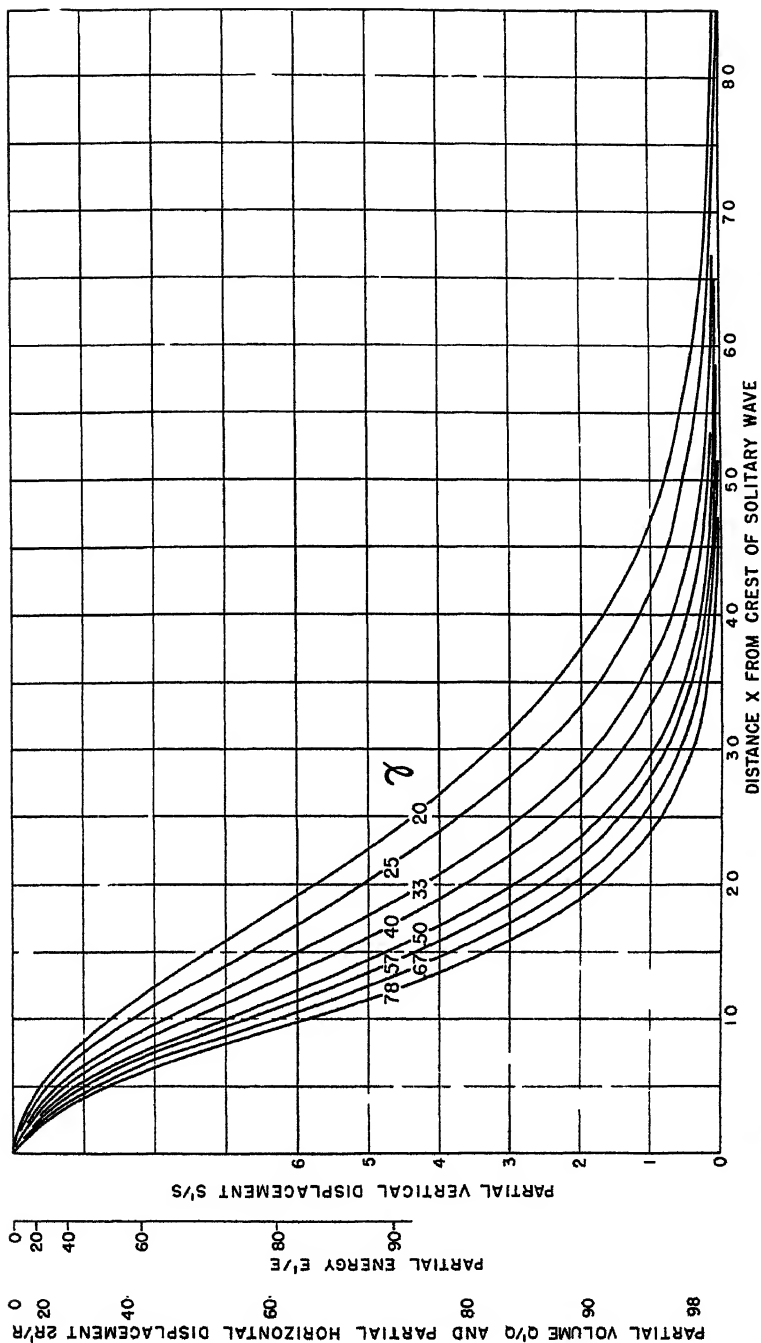


PLATE 2 Wave velocity (see EQUATION 6) and effective wave period (EQUATION 38) as functions of the depth h and wave height H . In order for the solitary wave theory to apply, the wave period should be larger than T_{eff} . The straight line marked $h = 1.28 H$ represents the extreme case of a breaking wave.



Decide use with distance from crest of partial energy volume, and particle displacements Energy and volume are concentrated within a narrow region to the east

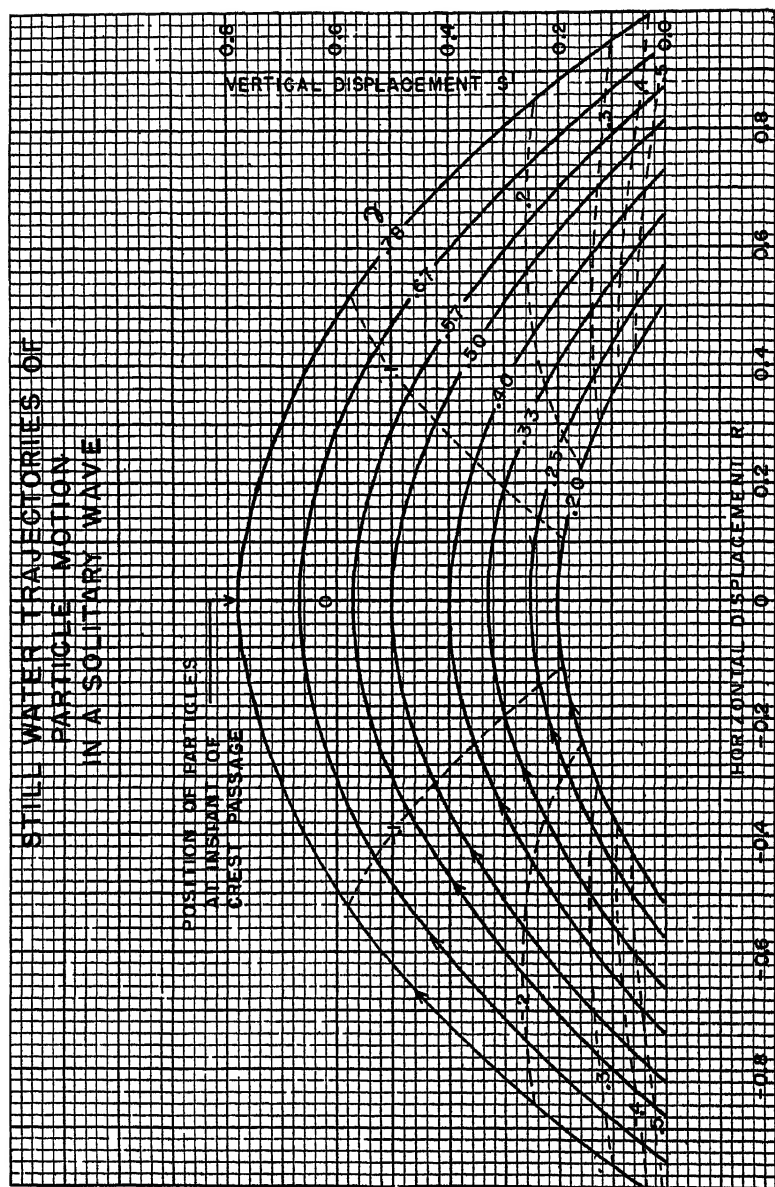
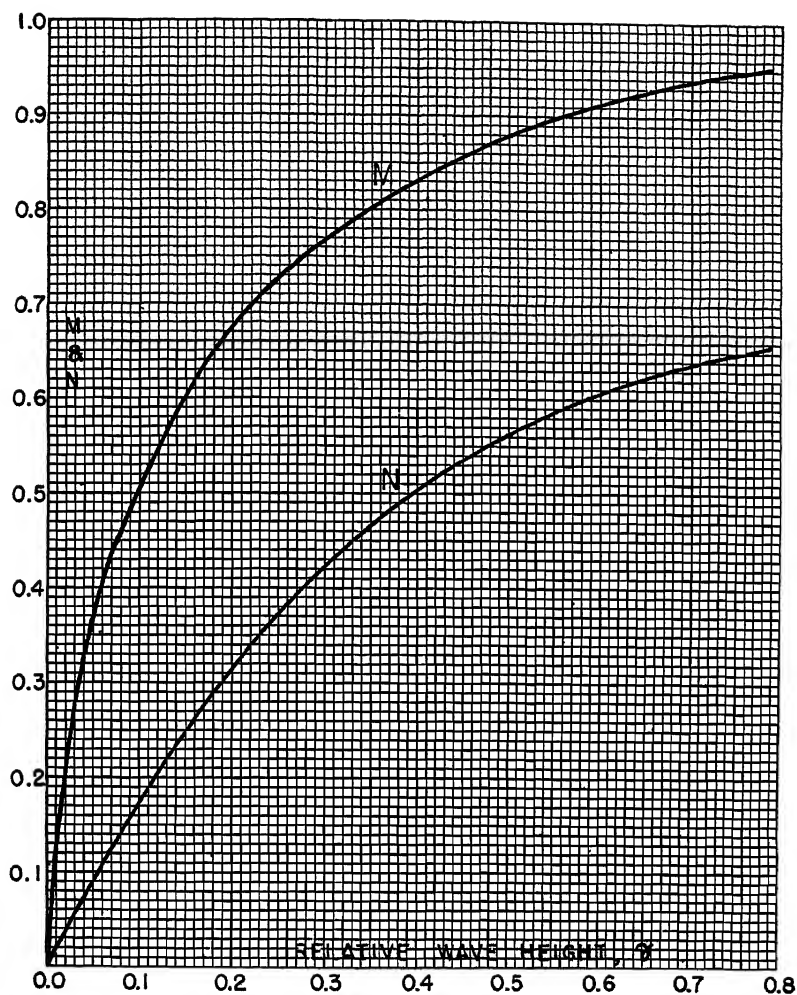


PLATE 4. The solid lines show the still water trajectories of a surface particle for various values of τ . Dashed lines give relative time $\tau C/k$ before (-) or after (+) crest passage.

PLATE 5. Values of parameters M and N defined by equations 27 and 28.

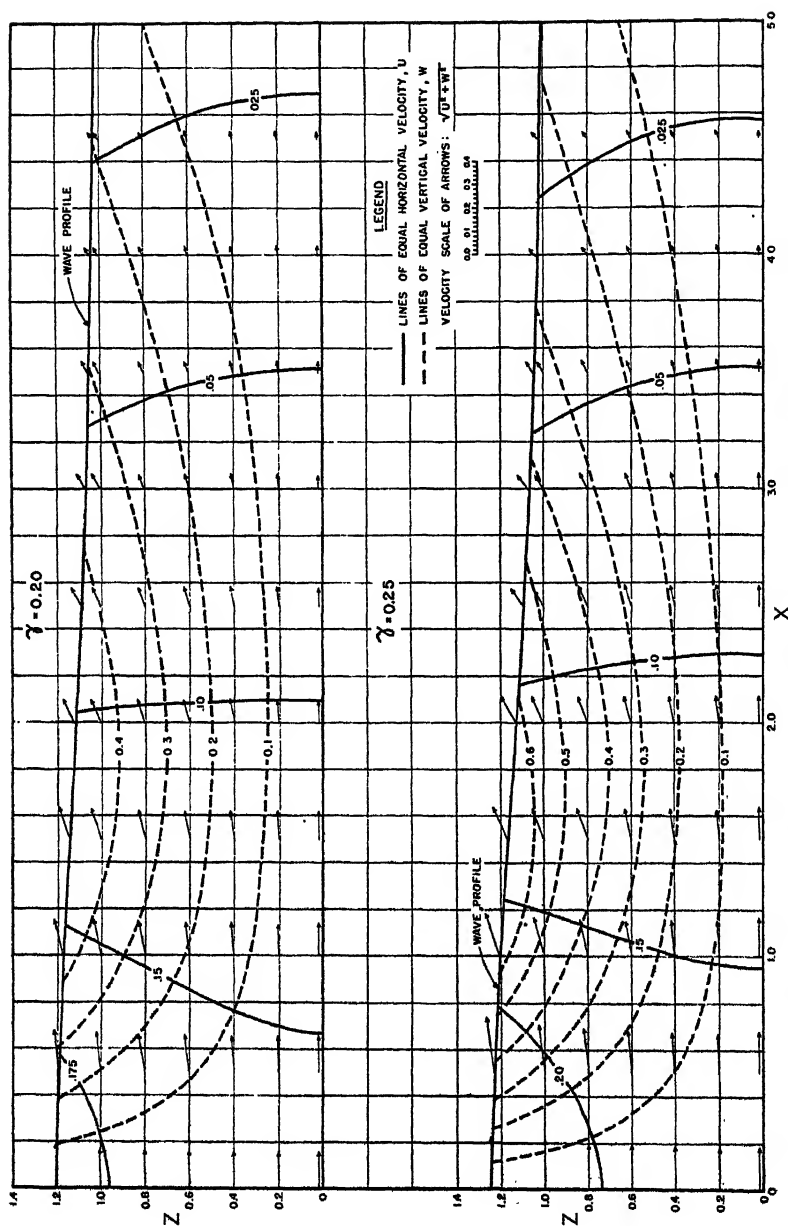


PLATE 6. Orbital motion in a solitary wave according to McCowan. Solid and dashed curves denote equal values of horizontal and vertical velocity, respectively. The arrows give the instantaneous direction of flow and their length is proportional to the orbital velocity according to the scale in the legend. Non-dimensional parameters are used throughout the plate.

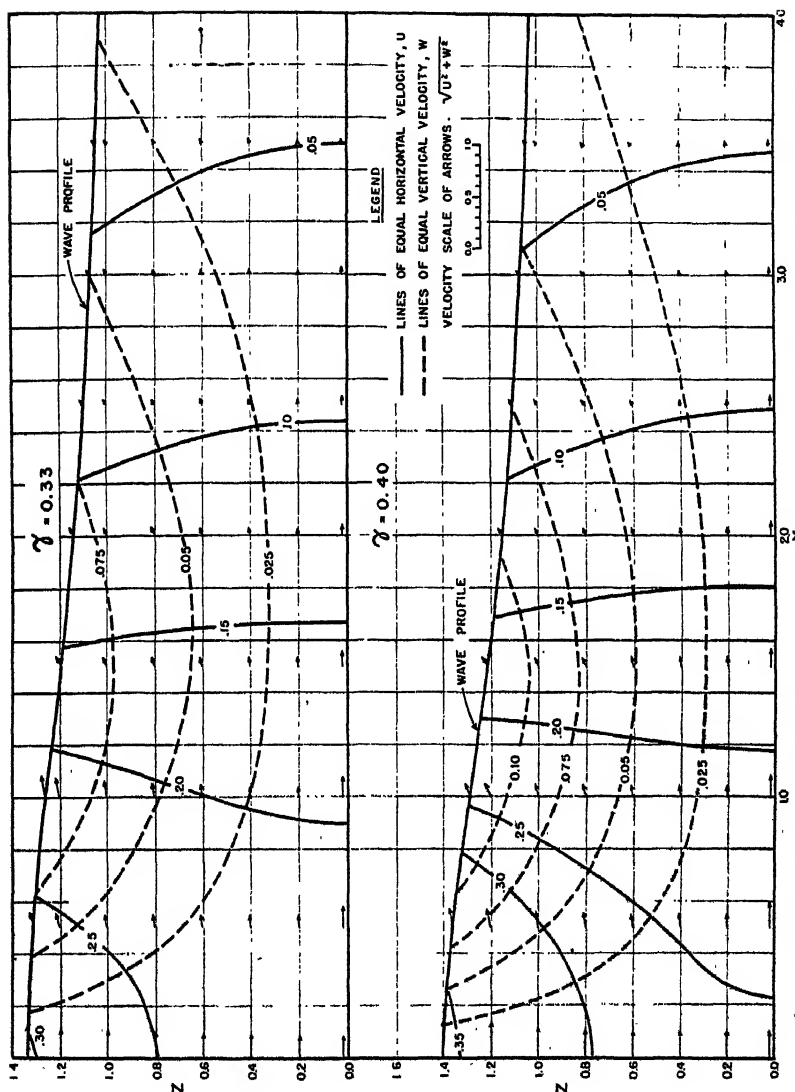


PLATE 7. Orbital motion in a solitary wave according to McCowan. Solid and dashed curves denote equal values of horizontal and vertical velocity, respectively. The arrows give the instantaneous direction of flow and their length is proportional to the orbital velocity according to the scale in the legend. Nondimensional parameters are used through the plate.

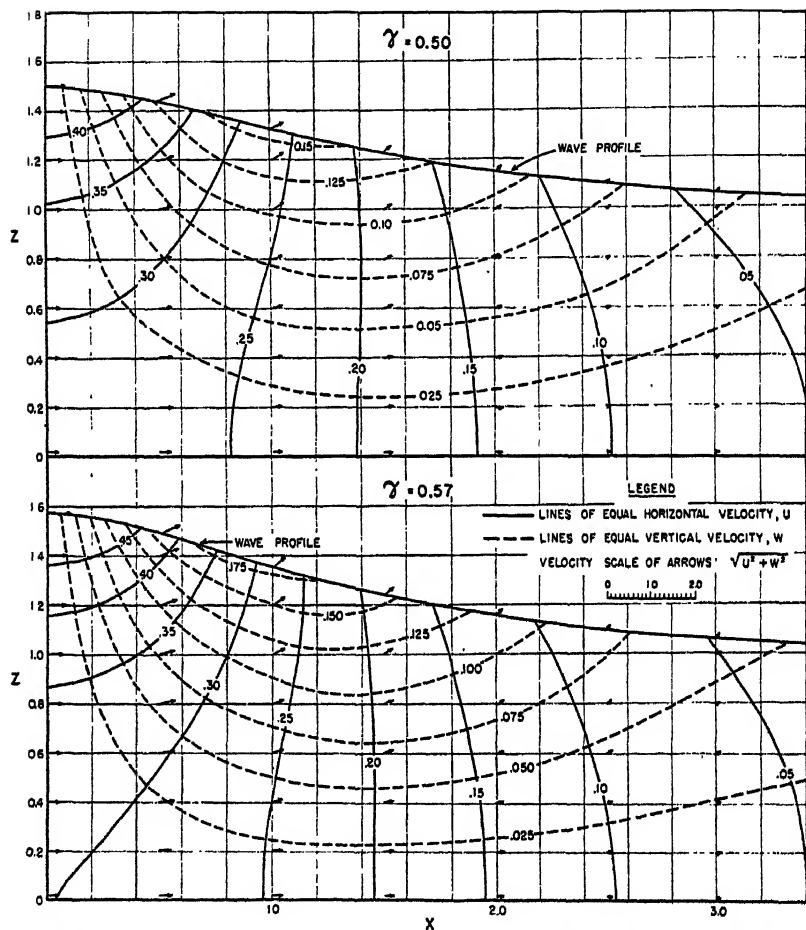


PLATE 8. Orbital motion in a solitary wave according to McCowan. Solid and dashed curves denote equal values of horizontal and vertical velocity, respectively. The arrows give the instantaneous direction of flow and their length is proportional to the orbital velocity according to the scale in the legend. Non-dimensional parameters are used throughout the plate.

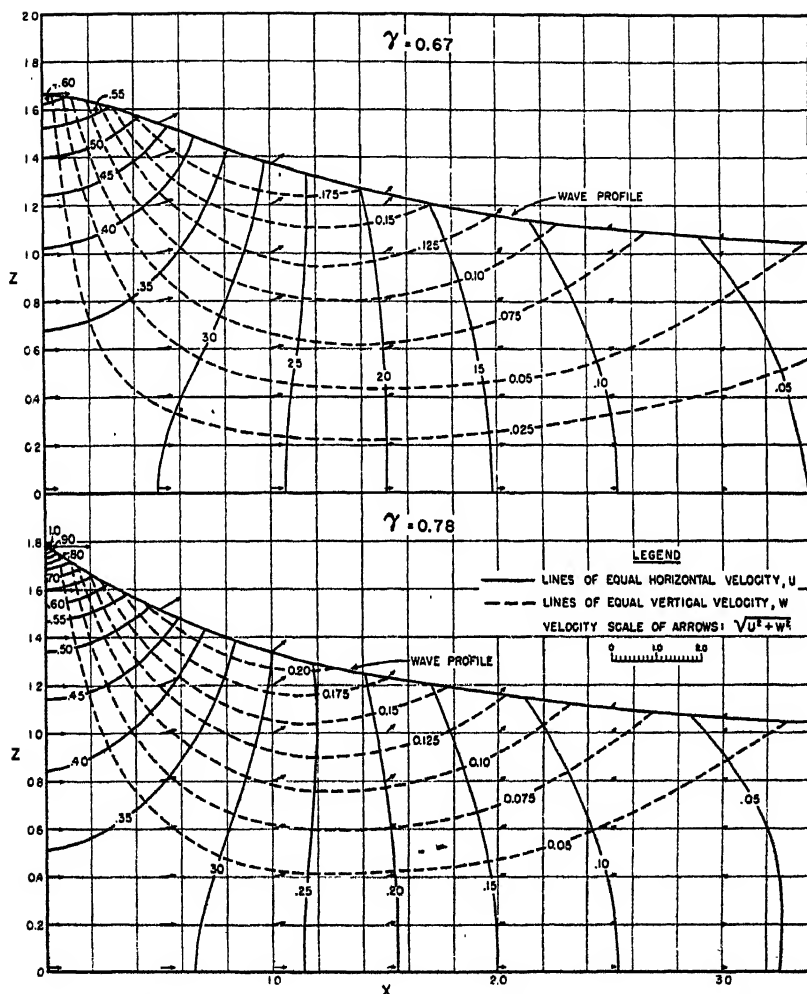


PLATE 9. Orbital motion in a solitary wave according to McCowan. Solid and dashed curves denote equal values of horizontal and vertical velocity, respectively. The arrows give the instantaneous direction of flow and their length is proportional to the orbital velocity according to the scale in the legend. Non-dimensional parameters are used throughout the plate. The lower figure represents the extreme case of a breaking wave.

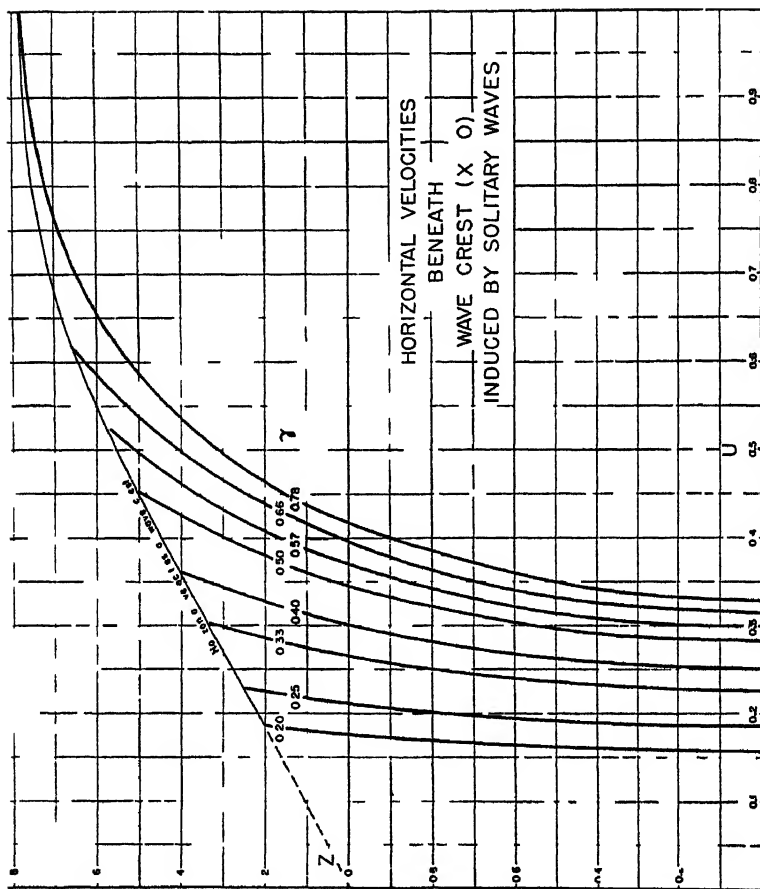


PLATE 10 Orbital velocity beneath crest according to EQUATION 30 The direction of orbital motion is horizontal and the velocity is at a maximum

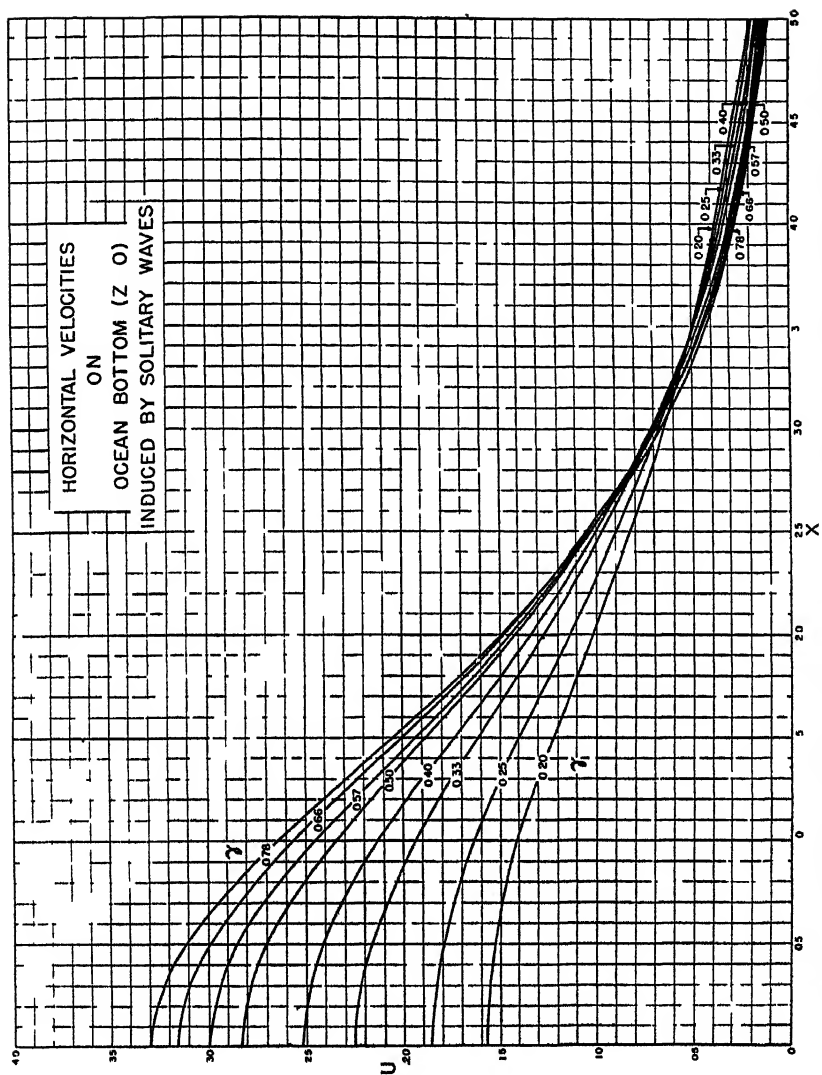


PLATE 11 Orbital velocity along the bottom according to EQUATION 31 The direction of orbital motion is horizontal

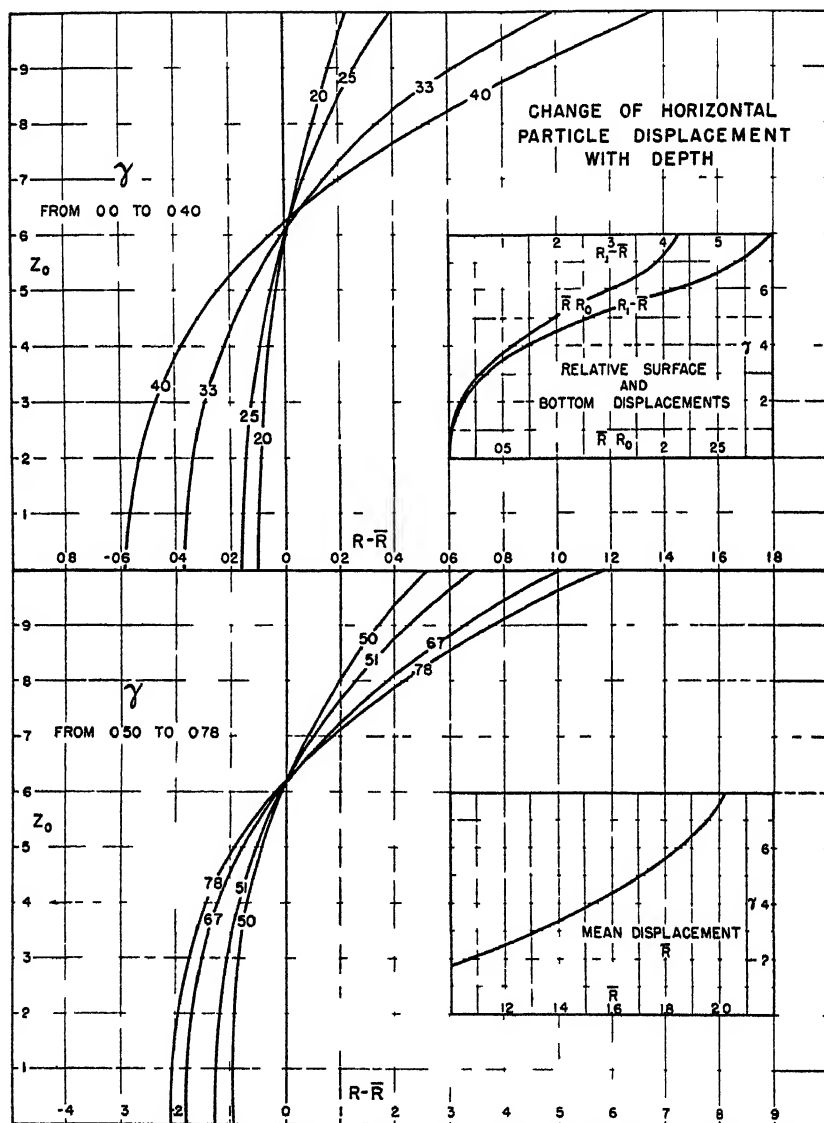


PLATE 12. Relative horizontal displacements (drift) during one wave period of particles initially at elevation Z_0 .

WAVE MOTION AT THE SURFACE OF A CURRENT WHICH HAS AN EXPONENTIAL DISTRIBUTION OF VORTICITY

By ABDUL JABBAR ABDULLAH

Massachusetts Institute of Technology, Cambridge, Massachusetts

Introduction. The study of waves that are propagated at the surface of a current is greatly complicated by different factors. Among those factors one might mention the earth's rotation, turbulence, viscosity, friction with outside bodies, and the vorticity distribution in the underlying basic current. Various attempts were made to investigate the influence of these factors. In these investigations, each factor might have been taken separately, or two or more of them might have been assumed to work together. The problem of the influence of vorticity distribution, however, has been restricted to some idealized pictures. The present paper might well be classified in this category. For in the present paper, the vorticity distribution is assumed to follow an exponential law. The vorticity is assumed to vanish at a great depth and to attain its maximum magnitude at the free surface. This, however, is not totally without a justifying reason. For, through the work of Ekman,¹ it is known that the surface oceanic currents follow, approximately, a spiral law. So that

$$\begin{aligned} U &= V_0 e^{\alpha z} \cos \left(\frac{\pi}{4} + \alpha z \right) \\ \text{and} \quad V &= V_0 e^{\alpha z} \sin \left(\frac{\pi}{4} + \alpha z \right) \end{aligned} \quad (1)$$

where U and V are the components of the velocity in the x - and y -directions respectively; the x - y plane being horizontal and in the undisturbed free surface, the z -axis being taken vertical (positive upwards). α is a constant which has the value

$$\alpha = + \sqrt{\frac{\rho \omega \sin \theta}{\mu}} \quad (2)$$

where ρ is the constant density, ω is the angular velocity of the earth's rotation, θ is the latitude, and μ is the coefficient of virtual viscosity. In EQUATION 1, V_0 is the absolute velocity of the water at the surface which has the value

$$V_0 = \frac{T}{\mu \alpha \sqrt{2}} \quad (3)$$

where T is the tangential pressure of the wind on the free surface. This tangential pressure is directed along the positive axis of y . The surface

velocity of the water, V_0 , is related to the wind velocity U' by the following equation, which was derived by Ekman (see Sverdrup²),

$$\frac{V_0}{U'} = \frac{0.0127}{\sqrt{\sin \theta}}. \quad (3.1)$$

Furthermore, it was pointed out by Ekman¹ that, according to EQUATION 1, the velocity vector turns around with depth, and at a certain depth D the direction becomes opposite to the direction at the surface. The depth D was called by Ekman the "Depth of Wind currents". D is related to α by the equation

$$D = \frac{\pi}{\alpha}. \quad (2.1)$$

Ekman's computations, however, show that by the time the depth D is reached the speed of the current is reduced to a small fraction of the speed at the surface.

In view of the preceding discussion and in view of the fact that waves do occur at the surface of ocean currents that are driven by winds, it seems interesting to investigate wave motion in these currents. The mathematical solution of this problem, in its general form, is rendered difficult because of several factors. First, the magnitude of the current speed decreases exponentially with depth, as is easily seen from EQUATION 1. This current distribution imposes a vorticity distribution. The subsequent wave motion thereby inherits a rotational motion, and the usual assumption of a velocity potential is therefore inadequate. A second factor is that the current changes direction. The motion, therefore, has to be treated in three dimensions. A third complicating factor is that with such a current distribution, like the one we are dealing with, the flow is not laminar. Turbulence becomes an important factor and it plays a role that may even overshadow that of gravity.

In the present paper, an attempt is made to deal with some aspects of this problem. To simplify matters only one of the above-mentioned factors is considered; namely, the effect of the speed distribution. The assumption is made that the effect of turbulence is to build up the basic current in a gradual manner. It is then assumed that the effect of turbulence on perturbation motion may be neglected. The flow may then be considered as laminar and free from viscosity. It is further assumed that the effect of the earth's rotation on the perturbation motion may also be neglected. Furthermore, the rotation of the velocity vector with depth is neglected and the current is assumed to be in the same direction throughout. The motion is therefore reduced to two dimensions.

The magnitude of the basic current, in the absence of vertical motion, is, from EQUATION 1,

$$U = V_0 e^{\alpha z}; \quad (4)$$

and, if the variation of the direction is neglected, the x -axis taken to coincide with the direction of the current, and the z -axis vertical and pointing upwards as before, the equations of the perturbation motion may be written as follows

$$\left. \begin{aligned} \left(\frac{\partial}{\partial t} + U \frac{\partial}{\partial x} \right) u + w \frac{\partial U}{\partial z} &= -\frac{1}{\rho} \frac{\partial p}{\partial x} \dots (a) \\ \text{and } \left(\frac{\partial}{\partial t} + U \frac{\partial}{\partial x} \right) w &= -\frac{1}{\rho} \frac{\partial p}{\partial z} \dots (b) \end{aligned} \right] \quad (5)$$

The linearized equation of continuity is

$$\frac{\partial u}{\partial x} + \frac{\partial w}{\partial z} = 0. \quad (6)$$

In EQUATIONS 5 and 6, u and w are the components of the perturbation velocity in the directions x and z respectively, p is the perturbation pressure, and ρ is the constant density.

The equation of continuity suggests the use of a stream function ψ such that

$$\left. \begin{aligned} u &= -\frac{\partial \psi}{\partial z} \\ \text{and } w &= \frac{\partial \psi}{\partial x} \end{aligned} \right] \quad (7)$$

Upon differentiating (5a) partially with respect to z and (5b) partially with respect to x and subtracting, the following equation is obtained

$$\left(\frac{\partial}{\partial t} + U \frac{\partial}{\partial x} \right) \xi + w \frac{\partial^2 U}{\partial z^2} = 0 \quad (8)$$

where $\xi = \frac{\partial u}{\partial z} - \frac{\partial w}{\partial x}$ is the vorticity. Upon substituting in (8) from (7) the former becomes

$$\left(\frac{\partial}{\partial x} + U \frac{\partial}{\partial t} \right) \nabla^2 \psi - \frac{\partial \psi}{\partial x} \frac{\partial^2 U}{\partial z^2} = 0 \quad (9)$$

where

$$\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial z^2}.$$

It is easily verified that EQUATION 9 agrees with the equation found by Rayleigh³ and later by G. I. Taylor.¹

The solution of (9) may be assumed to have the form

$$\psi = \varphi(x) e^{i(mx - \sigma t)} \quad (10)$$

where φ is a function of z only. Using the value of U as given in (4) and substituting from (10) in EQUATION 9, the following equation for φ is found

$$\varphi'' - \left(m^2 + \frac{\alpha^2 m U}{mU - \sigma} \right) \varphi = 0 \quad (11)$$

where the primes indicate differentiation with respect to z . The substitution

$$s = mU \quad (12)$$

transforms this equation to the following form

$$s^2 \frac{d^2 \varphi}{ds^2} + s \frac{d\varphi}{ds} - \left(\frac{m^2}{\alpha^2} + \frac{s}{s - \sigma} \right) \varphi = 0 \quad (13)$$

where now the independent variable is s . If, in EQUATION 13, the following substitution is made

$$\varphi = s^n v \quad (14)$$

where $n = \frac{m}{\alpha}$, the following equation for v is found

$$s \frac{d^2 v}{ds^2} + (2n + 1) \frac{dv}{ds} - \frac{v}{s - \sigma} = 0 \quad (15)$$

EQUATION 15 is a Frobenius type equation. It has singularities at $s = 0$, σ , $\pm \infty$. There are, therefore, three types of solutions. The solutions around $\pm \infty$ are, however, irrelevant to the physical problem which we are dealing with. This is because s is usually a small number and does not approach infinity. Hence we proceed to derive the other two solutions.

The solution around 0.

Upon setting

$$v = s^j \sum_{i=0}^{\infty} a_i s^i \quad (16)$$

and substituting in (15), the indicial equation is found to be

$$\sigma j(j + 2n) = 0. \quad (17)$$

Thus $j = 0$, or $-2n$. Taking $j = 0$ the general term is

$$a_r s^r = \frac{(\Gamma k)(k-1)(2k+1)(3k+5) \cdots (r-1)[(k+r-2)-1]}{\sigma a^r(r!) \Gamma(k+r-1)} a_0 \left(\frac{s}{\sigma} \right)^r \cdots \quad (18)$$

where Γ is the gamma function and $k = 2n + 1$, and the solution is

$$v_1 = \frac{1}{\sigma(2n+1)} \left[1 + \frac{2n}{2(n+1)} \left(\frac{s}{\sigma} \right) + \frac{n(4n+3)}{3!(n+1)(2n+3)} \left(\frac{s}{\sigma} \right)^2 + \dots \right] \quad (19)$$

and taking $j = -2n$, the solution may be written in the following concise form

$$v_2 = s^{-2n} K(s) \quad (20)$$

where $K(s)$ is an ascending power series in s . The explicit form of $K(s)$ need not be written here for reasons that will appear immediately. Hence the complete solution for EQUATION 13 is

$$\varphi = C_1 v_1 + C_2 v_2 \quad (21)$$

where C_2 and C_1 are arbitrary constants. The boundary condition that, at $z = -\infty$, the motion has to remain finite gives

$$C_2 = 0 \quad \text{and} \quad C_1 = C.$$

Hence the complete solution for our purpose is

$$\varphi = CS_1(s) \quad (22)$$

where

$$S_1(s) \equiv \frac{s^n}{\sigma(2n+1)} \left[1 + \frac{n}{2!(n+1)} \left(\frac{s}{\sigma} \right) + \frac{n(4n+3)}{3!(n+1)(2n+3)} \left(\frac{s}{\sigma} \right)^2 + \dots \right]. \quad (23)$$

From (10) it follows that

$$\psi = CS_1(s) e^{i(mx-\sigma t)} \quad (24)$$

and from (7) it follows that

$$\begin{aligned} w &= imCS_1(s) e^{i(mx-\sigma t)} \dots (a) \\ \text{and } u &= -\alpha CS_2(s) e^{i(mx-\sigma t)} \dots (b) \end{aligned} \quad (25)$$

where

$$S_2(s) \equiv \frac{s^n}{\sigma(2n+1)} \left[n + \frac{n(n+1)}{2!(n+1)} \left(\frac{s}{\sigma} \right) + \frac{n(n+2)(4n+3)}{3!(n+1)(2n+3)} \left(\frac{s}{\sigma} \right)^2 + \dots \right]. \quad (26)$$

To find p the assumption is made that

$$\frac{P}{\rho} = D(z)e^{i(mx-\sigma t)} \quad (27)$$

where $D(z)$ is a function of z only. If this value is substituted in (5), and the EQUATIONS 24, 25, and 26 are used, it follows that

$$\frac{P}{\rho} = - \left[\frac{\sigma - mU}{m} \alpha S_2(s) + \alpha U S_1(s) \right] C e^{i(mx-\sigma t)}. \quad (28)$$

The boundary condition at the free surface is (see Haurwitz⁴)

$$\left(\frac{\partial}{\partial t} + U \frac{\partial}{\partial x} \right) \frac{p}{\rho} - wg = 0; \quad \text{at } z = 0 \quad (29)$$

where g is the acceleration of gravity.

Now at $z = 0$

$$\begin{aligned} s_0 &= mV_0, \\ w_0 &= imCS_{01}(s_0) \end{aligned} \quad | \quad (30)$$

$$\text{and} \quad \frac{p_0}{\rho} = -(\epsilon - v_0)C\alpha S_{02} - C\alpha V_0 S_{01}$$

where the common exponential factor is understood. In (10), and also in (30), $m = 2\pi\lambda^{-1}$, where λ is the wave length, and σ is the frequency so that

is the phase velocity of the sinusoidal waves which travel at the surface. Upon substituting from (30) in (29), the following frequency equation is found

$$(\epsilon - V_0)^2 \frac{S_{02}}{S_{01}} + V_0(\epsilon - V_0) - \frac{g}{\alpha} = 0. \quad (31)$$

This equation can also be written in the form

$$\epsilon = V_0 - \frac{1}{2}V_0 \frac{S_{01}}{S_{02}} \pm \sqrt{\left(\frac{1}{2}V_0 \frac{S_{01}}{S_{02}} \right)^2 + \frac{g}{\alpha} \frac{S_{01}}{S_{02}}}. \quad (32)$$

This equation is not readily susceptible to computation. This is because both S_{01} and S_{02} involve the unknown quantity ϵ in the form of sums of terms in the ascending powers of this quantity. This may be seen readily from the identities (23) and (26). Nevertheless, a preliminary discussion of its implications may be given.

In the first place, it is easily seen from the identities (23) and (26) that the

limit of the ratio $\frac{S_{02}}{S_{01}}$ approaches n as s approaches zero. That is

$$\lim_{s \rightarrow 0} \left(\frac{S_{02}}{S_{01}} \right) \rightarrow n.$$

It is also obvious that s goes to zero if V_0 goes to zero. Thus

$$\lim_{V_0 \rightarrow 0} \left(\frac{S_{02}}{S_{01}} \right) \rightarrow n. \quad (33)$$

But, either from (31), or from (32), it is seen that

when $V_0 \rightarrow 0$

$$c \rightarrow \pm \sqrt{\frac{g}{\alpha} \lim_{V_0 \rightarrow 0} \left(\frac{S_{01}}{S_{02}} \right)} \quad (34)$$

or

$$c \rightarrow \pm \sqrt{\frac{g}{\alpha} \frac{1}{n}}$$

And since $n = \frac{m}{\alpha}$, (34) reduces to

$$c = \pm \sqrt{\frac{g}{m}}, \quad \text{when } V_0 = 0, \quad (35)$$

which is the Stokes's velocity (see Lamb).⁶ This result is in agreement with what is expected; since, in virtue of EQUATION 4, the basic current vanishes when the surface current vanishes. And if there is no basic current, the phase velocity follows the Stokes's formula for the first degree of approximation. It is also to be noticed from the identities (23) and (26) that the limit given in (33) is valid if V_0 is small compared with c . Thus, (35) can be put in the form

$$c \xrightarrow{(V_0/c) \rightarrow 0} \pm \sqrt{\frac{g}{m}}, \quad (36)$$

which shows that Stokes's formula is approached if the surface current is small compared with the wave velocity.

It appears from EQUATION 32 that there are two factors affecting the wave propagation in the case under consideration. These are gravity and the current shear. Because the shear is originally produced by turbulence, this latter factor will be called the turbulence factor. In EQUATION 32, the gravity factor is expressed by the term

$$\frac{g}{\alpha} \frac{S_{01}}{S_{02}},$$

which appears under the radical. The turbulence factor is expressed by the

term

$$\frac{1}{2}V_0 \frac{S_{01}}{S_{02}},$$

which appears both outside and inside the radical. Now, since the system of axes is so chosen as to make V_0 always positive, and since the ratio $\frac{S_{01}}{S_{02}}$ is always positive, it follows that

$$\frac{1}{2}V_0 \frac{S_{01}}{S_{02}} > 0.$$

Therefore

$$\left| \sqrt{\left(\frac{1}{2}V_0 \frac{S_{01}}{S_{02}}\right)^2 + \frac{g}{\alpha} \frac{S_{01}}{S_{02}}} \right| > \left| \frac{1}{2}V_0 \frac{S_{01}}{S_{02}} \right|. \quad (37)$$

This shows that nothing could be said about the relative magnitude of the speed of waves travelling in the direction of the current to the speed of waves travelling against the current. The two speeds may or may not be the same.

Again, if the absolute magnitude of the radical is taken, it is seen that

$$\left| \sqrt{\left(\frac{1}{2}V_0 \frac{S_{01}}{S_{02}}\right)^2 + \frac{g}{\alpha} \frac{S_{01}}{S_{02}}} - \left(\frac{1}{2}V_0 \frac{S_{01}}{S_{02}}\right) \right| < \left| \sqrt{\frac{g}{\alpha} \frac{S_{01}}{S_{02}}} \right|. \quad (38)$$

This shows that the total effect of turbulence and gravity is to give the waves a velocity which is smaller than that provided by gravity alone. Turbulence is, therefore, a retarding factor. This, again is expected, since turbulence is, in a way, a frictional force like viscosity.

It has already been remarked that neither EQUATION 32 nor EQUATION 31 is readily susceptible to computation. Therefore the numerical computations from these equations has to be done by some method of approximation. The method followed here is that of successive approximations. From the identities (23) and (26) it follows that the ratio $\frac{S_{02}}{S_{01}}$ has the value

$$\frac{S_{02}}{S_{01}} = n \left[1 + \frac{1}{2(n+1)} \left(\frac{V_0}{c} \right) + \frac{10n^2 + 19n + 3}{12(n+1)^2(2n+3)} \left(\frac{V_0}{c} \right)^2 + \dots \right] \quad (39)$$

and from EQUATION 31 it follows that

$$m = \frac{ng}{(c - V_0)^2 \frac{S_{02}}{S_{01}} + V_0(c - V_0)}. \quad (40)$$

The first approximation would be to put

$$\frac{S_{02}}{S_{01}} = n_1 = \frac{m_1}{\alpha}.$$

EQUATION 40 then gives

$$m_1 = \frac{g - \alpha V_0(c - V_0)}{(c - V_0)^2}. \quad (40.1)$$

This value for m may then be used in the EQUATION 39 and a second approximation may be found. The second approximation may again be used to find a third approximation, and so on.

It may be remarked that the series $\frac{S_{02}}{S_{01}}$ given in (39) converges rapidly

TABLE 1

$V_0 = 15$ cm/sec., $\alpha = 10^{-3}$ c.g.s., $g = 980$ c.g.s.

c m/ sec.	First approx.		Second approx.		Stokes values	
	m_1	λ ms.	m_2	λ m.	m_3	λ m.
1.00	$1.35 \cdot 10^{-1}$	0.465			1.355×10^{-1}	0.464
2	$2.86 \cdot 10^{-2}$	2.20			2.86×10^{-2}	2.19
3	$1.205 \cdot 10^{-2}$	5.12	$1.06 \cdot 10^{-2}$	5.92	1.206	5.21
4	0.657	9.55			0.661	9.50
5	0.414	15.18	0.423	14.81	0.416	15.11
6	$2.83 \cdot 10^{-3}$	22.09			2.85×10^{-3}	22.01
7	2.07	33.20			2.09	30.05
8	1.565	40.02			1.585	39.7
9	1.235	50.08			1.25	50.03
10	0.994	63.10	$0.99 \cdot 10^{-3}$	63.4	1.01	62.02
11	$8.85 \cdot 10^{-4}$	71.00			8.97×10^{-4}	69.9
12	6.82	92.05			6.95	90.4
13	5.82	107.80			5.93	106.0
14	5.01	125.10			5.10	123.3
15	4.32	145.20	$4.3 \cdot 10^{-4}$	146.0	4.42	142.0
16	3.81×10^{-4}	164.90			3.905×10^{-4}	160.8
17	3.36	187.00			3.45	182.0
18	3.02	208.60			3.08	202.0
19	2.68	234.10			2.76	227.8
20	2.41	260.50	$2.41 \cdot 10^{-4}$	260.50	2.48	253.0
25	1.54	407.00			1.605	391
30	1.05	598.00			1.1	571
35	0.761	824.00			0.804	781
40	0.572	1099.00			0.616	1021

and uniformly for values of $\frac{V_0}{c}$ that are less than unity. The fact that the series converges rapidly enables us to neglect terms of third and higher order, and thus the numerical computation is made easier.

TABLES 1 and 2 give the first and second approximations as computed by the method of successive approximations.

From EQUATION 3.1 it can be seen that a surface velocity of 15 cm/sec is caused by a wind of about 10 m/sec, at latitude 43° . The value of α was computed from the corresponding value of D as given by Sverdrup³ (page 23). The Stokes's values were computed from the equation

$$m = \frac{g}{(c - V_0)^2}.$$

FIGURE 1 is a plot of these values.

It can be seen from these tables that the second approximation gives for m values that are slightly higher than those given by the first approximation. The difference, however, is small and the correction involved is of the order of 3 per thousand.

FIGURE 2 is a plot based on TABLE 2.

It can also be seen either from FIGURES 1 and 2 or from TABLES 1 and 2 that Stokes's values for the velocity are higher than the values given by (40). This is in agreement with the discussion given in this paper.

TABLE 2
 $V_0 = 10 \text{ cm/sec.}, \alpha = 10^{-4} \text{ c.g.s.}, g = 980 \text{ c.g.s.}$

$c \text{ m/}$ sec.	First approx.		Second approx.		Stokes values	
	m_1	$\lambda \text{ ms.}$	m_2	$\lambda \text{ m.}$	m_3	$\lambda_3 \text{ m.}$
1	1.555	0	∞	0	∞	0.0
2	2.555×10^{-2}	0.641			9.8×10^{-2}	1.00
3	2.555	2.562			2.45	2.64
4	0.586	5.78			1.09	5.56
5	0.50×10^{-3}	10.30	6.08×10^{-3}	10.42	0.61×10^{-3}	10.75
6	2.59	16.15			3.92	16.28
7	2.405	23.09			2.72	23.00
8	1.485	31.60			2.0	31.42
9	1.513	41.04			1.53	41.00
10	1.20	52.30	1.18×10^{-3}	53.60	1.21	51.80
11	9.7×10^{-4}	64.80			0.98	64.00
12	8.0	78.50			0.81	77.25
13	6.22	100.80			0.68	92.50
14	5.72	109.90			0.58	108.20
15	4.92	127.60	4.91×10^{-4}	127.80	5.0×10^{-4}	125.50
16	4.34	144.90			4.35	144.10
17	3.77	167.00			3.83	163.80
18	3.33	188.50			3.39	185.00
19	2.97	210.80			3.02	207.90
20	2.66	236.00	2.63×10^{-4}	238.5	2.70	232.30
25	1.66	378.00	1.65	380.4	1.71	367.8
30	1.13	561.00	1.12	560.8	1.66	541.0
35	0.8	764.00			0.848	742.1
40	0.618	1015.00			0.643	979.0

The solution around σ .

The previous discussion was based on the assumption that s is very small; that is, that the ratio $\frac{V_0}{c}$ is in the neighborhood of zero. It is thought that this has some physical significance, since the wave velocity is usually much greater than the velocity of the basic current. However, a solution for EQUATION 15 may be found for values of s that are in the neighborhood of σ ; that is, for values of $\frac{V_0}{c}$ near unity. This does not seem to have much physical bearing since c is usually much greater than V_0 . Nevertheless, the solution for this case will here be given for the sake of completeness.

If in EQUATION 15 the substitution

$$s - \sigma = s_1$$

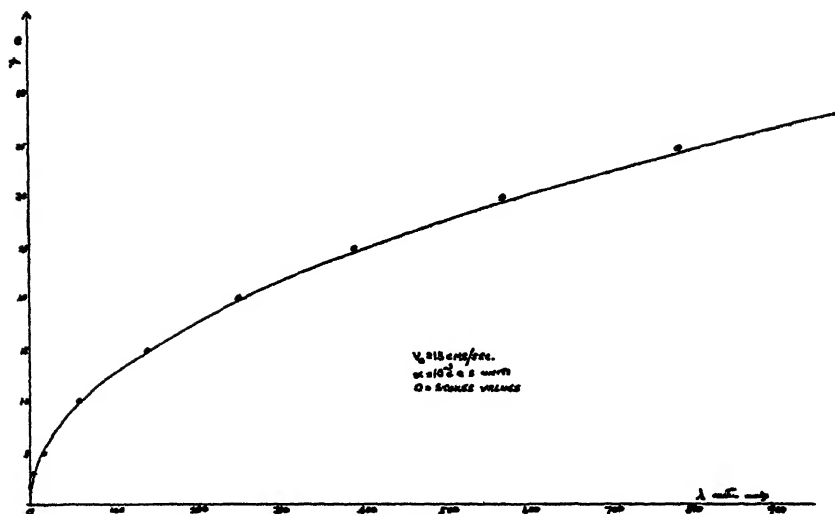


FIGURE 1.

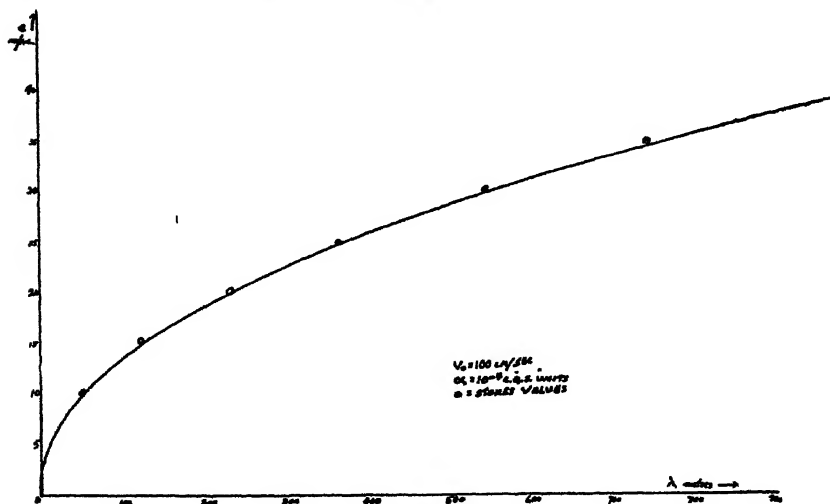


FIGURE 2.

is made, the equation takes the form

$$(s_1 + \sigma)s_1 \frac{d^2 v}{ds_1^2} + ks_1 \frac{dv}{ds_1} - v = 0, \quad k = 2n + 1. \quad (41)$$

The solution for (41) may be assumed to have the form

$$v = s_1^k (a_0 + a_1 s_1 + a_2 s_1^2 + \cdots a_n s_1^n + \cdots).$$

The indicial equation is

$$j(j-1) = 0$$

and the $r+1$ th term is

$$a_{r+1}s_1^{r+1} = \frac{1 - (r+j)(r+j-1-k)}{\sigma(r+j)(r+j+1)} a_r s^r.$$

Taking $j = 0$, this becomes

$$a_{r+1}s^{r+1} = \frac{1 - r(r-k-1)}{\sigma r(r+1)} a_r s^r$$

and the series solution in terms of the original variable s is

$$v = C(s - \sigma)H_1 \quad (42)$$

where

$$H_1 = \left[1 + \frac{1+k}{2} \left(\frac{s}{\sigma} - 1 \right) + \frac{(1+k)(2k-1)}{2^2 \cdot 3} \left(\frac{s}{\sigma} - 1 \right)^2 + \dots \right]. \quad (43)$$

Taking $j = 1$, a solution identical to (43) is found. An additional logarithmic solution may then be written, but this is not necessary because this additional solution will vanish in virtue of the lower boundary condition. Hence the complete solution for our purpose is

$$\left. \begin{aligned} \psi &= Cs^n(s - \sigma)H_1 e^{i(mx - \sigma t)} \\ \text{Hence } u &= imCs^n(s - \sigma)H_1 e^{i(mx - \sigma t)} \\ \text{and } u &= [\alpha Cn(s - \sigma)s^n H_1 + \alpha Cs^{n+1} H_2] e^{i(mx - \sigma t)} \end{aligned} \right] \quad (44)$$

where

$$H_2 = 1 + \frac{2(1+k)}{2\sigma} (s - \sigma) + 3 \frac{(1+k)(2k-1)}{2^2 \cdot 3 \cdot \sigma^2} (s - \sigma)^2 + \dots \quad (45)$$

And from (5) it is found that

$$\frac{P}{\rho} = \left\{ \alpha \frac{\sigma - s}{m} s^n [n(s - \sigma)H_1 + sH_2] - \frac{\alpha s^{n+1}}{m} (s - \sigma)H_1 \right\} C e^{i(mx - \sigma t)}. \quad (46)$$

Upon substituting in the boundary condition for the upper free surface, which is given in EQUATION 29, the following frequency equation is found

$$\sigma - mV_0 = \frac{\alpha V_0}{2} \left(\frac{H_2}{H_1} - 1 \right) \pm \sqrt{\left(\frac{\alpha V_0}{2} \right)^2 \left(\frac{H_2}{H_1} - 1 \right)^2 + mg}. \quad (47)$$

It can be seen that, in (47), when

$$V_0 \rightarrow 0$$

$$\sigma \rightarrow \pm \sqrt{mg}.$$

Also when $V_0 \rightarrow c$, the ratio $\frac{H_2}{H_1} \rightarrow 1$, and

$$\sigma \rightarrow \pm \sqrt{mg}.$$

which is the value that one expects.

Stability of Surface Waves.

To investigate the stability of the waves under consideration it will be assumed that, in the undisturbed case, the air flows with a velocity which is constant with height. Let the component of this velocity in the direction of x be U' . It will further be assumed that the wave motion in the air is symmetrical with respect to the y -axis. The air will be considered as a homogeneous non-viscous fluid of constant density ρ' . The compressibility of the air is neglected because the speeds we are dealing with are much smaller than the speed of sound. Under these assumptions the equations of the perturbation motion in the air will be

$$\left. \begin{aligned} \left(\frac{\partial}{\partial t} + U' \frac{\partial}{\partial x} \right) u' &= -\frac{1}{\rho'} \frac{\partial p'}{\partial x} \quad (a) \\ \text{and } \left(\frac{\partial}{\partial t} + U' \frac{\partial}{\partial x} \right) w' &= -\frac{1}{\rho'} \frac{\partial p'}{\partial z} \quad (b) \end{aligned} \right] \quad (48)$$

and the linearized equation of continuity is

$$\frac{\partial u'}{\partial x} + \frac{\partial w'}{\partial z} = 0 \quad (49)$$

where the primes refer to the upper fluid, which is the air.

Upon differentiating (48a) partially, with respect to z , and (48b), with respect to x , and subtracting, the following equation is obtained

$$\left(\frac{\partial}{\partial t} + U' \frac{\partial}{\partial x} \right) \left(\frac{\partial w'}{\partial x} - \frac{\partial u'}{\partial z} \right) = 0. \quad (48.1)$$

EQUATION 49 suggests the use of the stream function ψ' , such that

$$\left. \begin{aligned} u' &= -\frac{\partial \psi'}{\partial z} \\ w' &= \frac{\partial \psi'}{\partial x} \end{aligned} \right] \quad (50)$$

Upon substituting from (50) in (48.1), the latter becomes

$$\left(\frac{\partial}{\partial t} + U' \frac{\partial}{\partial x} \right) \nabla^2 \psi' = 0 \quad (51)$$

$$\text{where } \nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial z^2}.$$

Let the solution of ψ' be of the form

$$\psi' = \varphi'(z)e^{i(mx-\sigma t)}$$

where $\varphi'(z)$ is a function of z only. Upon substituting in (51), the following differential equation for φ' is obtained

$$\frac{d^2 \varphi'}{dz^2} - m^2 \varphi' = 0.$$

This has the following solution

$$\varphi' = C_1 e^{mz} + C_2 e^{-mz}.$$

Because φ' has to remain finite when z becomes infinite, it follows that

$$C_1 = 0, \quad C_2 = C'$$

hence

$$\varphi' = C' e^{-mz}$$

and therefore

$$\psi' = C' e^{-mz} e^{i(mx-\sigma t)} \quad (52)$$

and, from (50),

$$u' = C' m e^{-mz} e^{i(mx-\sigma t)} \quad (a)$$

and

$$w' = i C' m e^{-mz} e^{i(mx-\sigma t)}. \quad (b) \quad (53)$$

Also, from (48), it is found that

$$P' = C' \rho' (\sigma - m U') e^{-mz} e^{i(mx-\sigma t)}. \quad (c)$$

The boundary conditions at the surface which separates the upper fluid (the air) from the lower fluid (the water) is

$$P + p - (P' + p') = 0 \quad (54)$$

where the capitals stand for the undisturbed conditions and the small letters for the perturbation quantities. Under the previous assumptions (54) may be written in the following differential form

$$\frac{\partial}{\partial t} (p - p') + \left[\frac{U}{U'} \right] \frac{\partial}{\partial x} (p - p') + \left[\frac{w}{w'} \right] \frac{\partial}{\partial z} (P - P') = 0, \quad (55)$$

at $z = 0$.

Because both fluids are assumed to have no vertical accelerations in the

undisturbed case, the undisturbed pressures are given by the hydrostatic equation, and therefore

$$\frac{\partial}{\partial z} (P - P') = -g(\rho - \rho'). \quad (56)$$

Upon substituting in (55) from (30), (53), and (56), first for the upper fluid then for the lower, the following two equations are obtained

$$\begin{aligned} &(\sigma - mU')\alpha\rho[(c - V_0)S_{02} + V_0S_{01}]C \\ &\quad + [(\sigma - mU')\rho'm(c - U') - g(\rho - \rho')m]C' = 0 \\ &\{(\sigma - mV_0)\alpha\rho[(c - V_0)S_{02} + V_0S_{01}] \\ &\quad - mS_{01}g(\rho - \rho')\}C + (\sigma - mV_0)\rho'm(c - U')C' = 0 \end{aligned}$$

where $c = \frac{\sigma}{m}$ is the phase velocity of waves; and, upon equating the ratio of the constants, the following frequency equation is found

$$m\rho'(c - U')^2 + \alpha\rho(c - V_0)^2 \frac{S_{02}}{S_{01}} + \alpha\rho\rho'(c - V_0) - g(\rho - \rho') = 0. \quad (57)$$

Solving this equation for c , the following is found

$$\begin{aligned} c = & \frac{\rho'U' + \frac{\alpha S_{02}}{mS_{01}}\rho V_0 - \frac{1}{2}\frac{\alpha}{m}\rho V_0}{\rho' + \frac{\alpha S_{02}}{mS_{01}}\rho} \\ & \pm \sqrt{\frac{g(\rho - \rho')}{m\left(\rho' + \frac{\alpha S_{02}}{mS_{01}}\rho\right)} - \frac{m\alpha \frac{S_{02}}{S_{01}}\rho\rho'(U' - V_0)^2 + m\alpha\rho\rho'V_0(U' - V_0) - (\frac{1}{2}\alpha\rho V_0)^2}{\left(m\rho' + \alpha \frac{S_{02}}{S_{01}}\rho\right)^2}} \end{aligned} \quad (58)$$

It may be remarked that, if in this equation ρ' is set equal to zero, the equation reduces to

$$c = V_0 - \left(\frac{1}{2}V_0 \frac{S_{01}}{S_{02}}\right) \pm \sqrt{\frac{g}{\alpha} \frac{S_{01}}{S_{02}} + \left(\frac{1}{2}V_0 \frac{S_{01}}{S_{02}}\right)^2}, \quad (32.1)$$

which is the same as EQUATION 32. Furthermore, if the lower fluid is assumed to have a constant basic current, i.e. if $\alpha \rightarrow 0$, then $\frac{S_{02}}{S_{01}}\alpha \rightarrow m$ and the equation reduces to

$$\frac{\rho'U' + \rho V_0}{\rho' + \rho} \pm \sqrt{\frac{g}{m} \frac{\rho - \rho'}{\rho + \rho'} - \frac{\rho\rho'(U' - V_0)^2}{(\rho + \rho')^2}}, \quad (58.2)$$

which is the well known equation for the velocity of waves at the inner surface separating two incompressible fluids (see Lamb⁶ p. 373).

It can be seen from EQUATION 58 that the stability of the waves depends upon the value of the radical. The waves will be stable if

$$\frac{g}{m} \frac{(\rho - \rho') \left(m\rho' + \alpha \frac{S_{02}}{S_{01}} \rho \right)}{\rho\rho'} \geq \left[\alpha \frac{S_{02}}{S_{01}} (\Delta U')^2 + \alpha V_0 (\Delta U) - \left(\frac{1}{2} \frac{\alpha^2}{m} \frac{\rho}{\rho'} V_0^2 \right) \right] \quad (59)$$

where

$$\Delta U' = U' - V_0.$$

Or if

$$\lambda \geq \frac{2\pi}{g} \frac{\frac{\alpha}{m} \frac{S_{02}}{S_{01}} \rho\rho' (\Delta U')^2 + \frac{\alpha}{m} \rho\rho' V_0 (\Delta U) - \left(\frac{1}{2} \frac{\alpha^2}{m} \frac{\rho}{\rho'} V_0^2 \right)}{(\rho - \rho') \left(\rho' + \frac{\alpha S_{02}}{m S_{01}} \rho \right)} \quad (60)$$

where λ is the wave length.

The inequality (60) is not readily susceptible to computations, because of the appearance of the unknown m in the right hand side, and because the ratio $\frac{S_{02}}{S_{01}}$ involves the same quantity in the form of a sum of terms in the ascending powers of m . A certain amount of discussion is, however, possible. EQUATION 60 may be compared with the corresponding equation which is derived from (58.2). There, it is found that the condition for stability is given by the inequality

$$\lambda_s \geq \frac{2\pi}{g} \frac{\rho\rho' (\Delta U')^2}{(\rho - \rho')(\rho + \rho')} \quad (61)$$

Since $\frac{\alpha}{m} \frac{S_{02}}{S_{01}}$ is not very different from unity, the denominator of both inequalities are nearly equal. The first term on the right hand side of (60) is therefore nearly equal to the right hand side of (61). Hence, for a first approximation the inequality (60) can be written in the following form

$$\lambda \geq \lambda_s + \frac{\frac{\alpha}{m} \rho\rho' V_0 (\Delta U) - \left(\frac{1}{2} \frac{\alpha^2}{m} \frac{\rho}{\rho'} V_0^2 \right)}{(\rho - \rho')(\rho + \rho')} \quad (60.1)$$

(60.1) gives for the critical wave length a value which is somewhat higher than that computed on the basis of the simpler inequality (61). Thus, for $U' = 10$ m/sec, $\rho = 1$, $\rho = 1.3 \times 10^{-3}$ and $\alpha = 10^{-3}$ c.g.s. units, the critical

wave length for $V = 15$ cm/sec is 8.124 cms, and for $V = 10^2$ cm/sec it is 9.683 cms, whereas (61) gives 8.1 cm and 6.25 cm respectively. The correction, therefore, becomes more pronounced as the surface current increases.

Summary

In this paper, a first attempt is made towards a solution of wave motion superposed on a current which is caused by winds. The current is assumed to follow the Ekman's spiral in speed but to remain constant in direction. Frequency equations are derived in the form of series solutions, and approximate computations on the basis of these series are made. The results are compared with gravitational waves that are travelling on a constant current. The stability of surface waves is also discussed. The assumption is made that the wind velocity is constant with height.

References Cited

1. EKMAN, V. W. 1905-1906. On the influence of the earth rotation on ocean currents. Arkiv Math., Astron. + Phys. B 2 (11). Stockholm.
2. SVERDRUP, H. U. 1943. *Oceanography For Meteorologists*. Prentice-Hall, Inc. New York.
3. RAYLEIGH, J. W. 1916. Phil. Mag. London, 32: 529.
4. TAYLOR, G. I. 1931. Proc. R. Soc. 132 (Ser. A): 499.
5. HAURWITZ, B. 1941. *Dynamic Meteorology*. McGraw-Hill. New York.
6. LAMB, H. 1945. *Hydrodynamics*, 6th edition. Dover Publication. New York.

DETERMINATION OF THE DEPTH OF AN UNDERWATER EXPLOSION FROM MEASUREMENTS OF THE DOME OF SPRAY*

By CHAIM L. PEKERIS

Department of Applied Mathematics, The Weizmann Institute, Rehovoth, Israel and the Institute for Advanced Study, Princeton, N. J.

With a view to a possible application to the interpretation of artificially generated surface waves under controlled experimental conditions, an account is given here of some observations made in 1944 of the dome of spray which is thrown up by a statically-fired, underwater charge. Motion pictures were obtained of the rise of the dome of spray when charges of known weight and depth of suspension were set off, and methods were developed for determining the depth of the charges from the photographs alone. This study proved that the initial spray velocity is equal to the acoustic particle velocity imparted by the shock wave, an assumption which was first made by Shaw.

According to Shaw's theory, based on the above assumption, the distance out from the center of the dome to a radius where the dome-height has decreased to half the central height is equal to the depth of the explosion. An analysis of the dome shapes of some fifty static shots fired mainly at depths of 30 to 40 feet, by a modification of Shaw's method, gave values which were within about 10 per cent of the actual depths. This result confirms the applicability of Shaw's method to statically-fired charges.

An alternative method for determining the depth of an underwater explosion was also developed, based on measurements of the initial rate of rise of the center of the dome of spray. The results of analysis of the 50 shots by this method show that the accuracy obtainable is around 6 per cent.

When the distance of camera to charge is not known, neither of the above methods is applicable. But the two methods may then be combined to obtain a value for the ratio of velocity at the center of the dome to the depth of charge, from which the depth can then be inferred. This method is accurate to about 5 per cent.

The interval between the first appearance of the dome and the time when the plumes emerge above the dome was found to be highly variable for charges at the same depth. It is, therefore, not believed that the depth of an underwater explosion can be inferred from this time interval alone.

Shaw's method of determining the depth of an underwater explosion from the shape of the dome of spray. The idea underlying Shaw's method is that the initial rate of rise of the spray is equal to the acoustic velocity imparted to particles in the surficial layer by the peak of the shock wave. If P denotes

* Based on work done at Columbia University under contract with O.S.R.D.

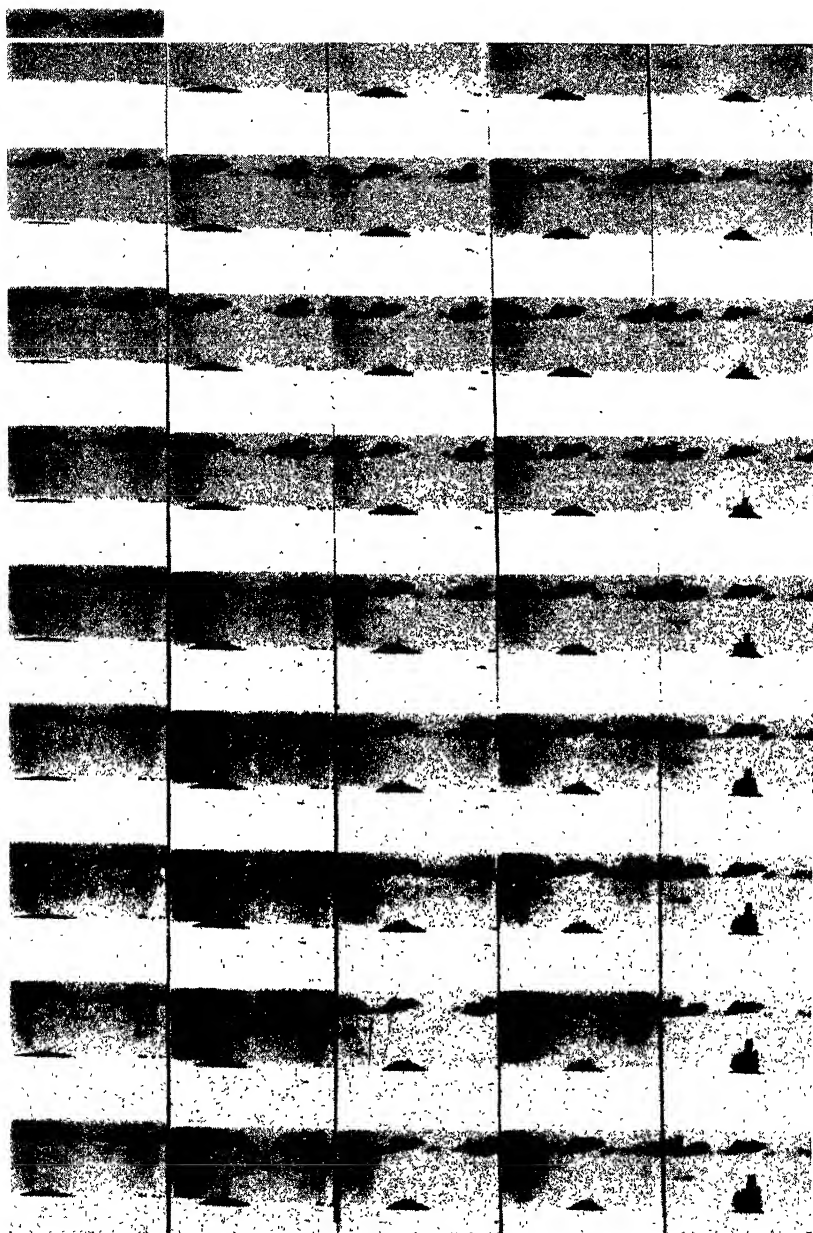


FIGURE 1. The dome of spray due to an underwater explosion.

the peak pressure in the shock wave, ρ and c the density and sound velocity of the water, then the particle-velocity v is to a sufficient approximation

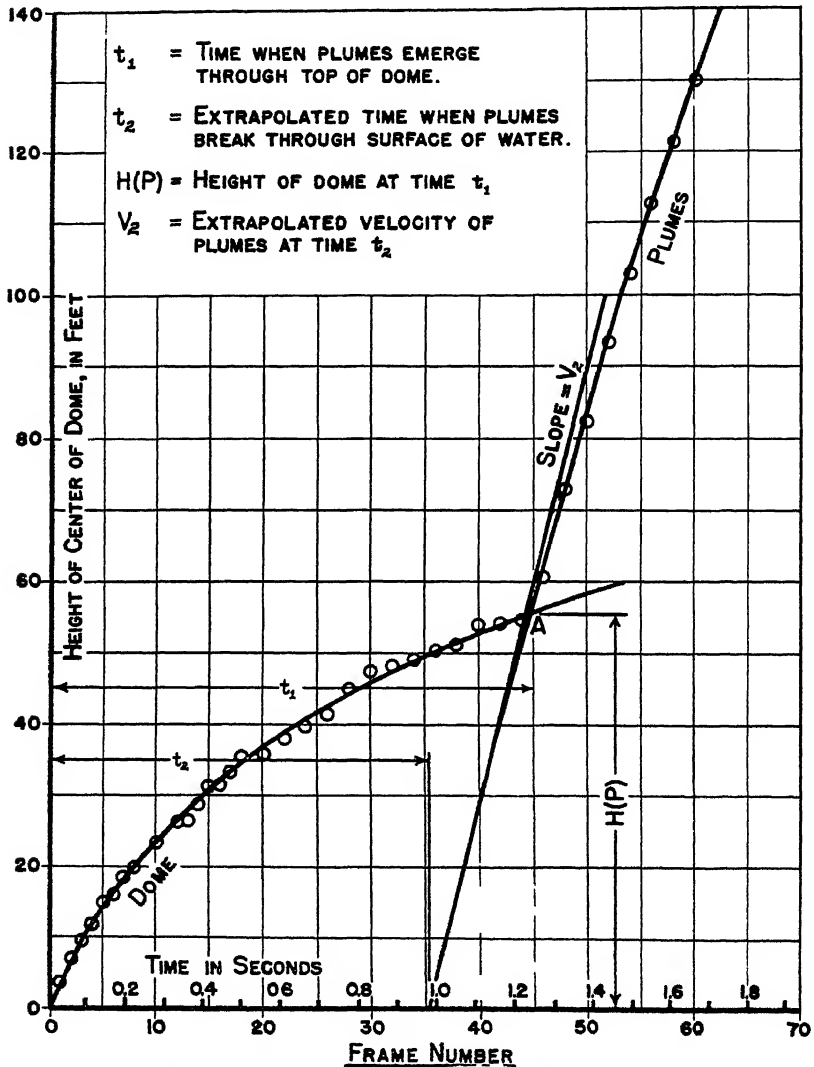


FIGURE 2. Height of center of dome of spray as a function of time.

(see discussion below) given by $v = P/\rho c$. Referring to FIGURE 3, the velocity imparted to a particle at O in the surface by the direct pressure wave from the charge at C is in the direction CO . The image source imparts to it a velocity of equal magnitude in the direction OI . In the resulting velocity, V , the horizontal components cancel out and

$$V = 2v \cos \theta = \frac{2P}{\rho c} \cos \theta = \frac{2PD}{\rho c R}. \quad (1)$$

The variation of peak pressure P with distance from the charge has been determined empirically for a number of explosives.

In the range of values of charge weight W and distance R from charge which is pertinent to our application, the relation is

$$P = A(W^{1/3}/R)^n, \quad (2)$$

where A and n depend on the kind of explosive. The value of n is greater than 1 on account of dissipation of the shock front. Combining (1) and (2) we obtain for the variation of initial velocity with horizontal distance r

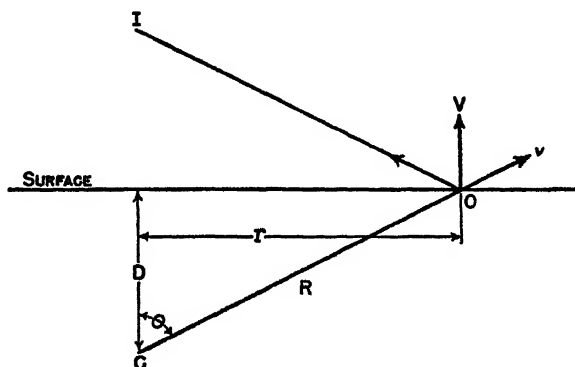


FIGURE 3.

along the surface,

$$V(r) = \frac{2ADW^{n/3}}{\rho c(r^2 + D^2)^{(1+n)/2}}. \quad (3)$$

$$\frac{V(r)}{V(o)} = \frac{1}{(1 + r^2/D^2)^{(1+n)/2}}. \quad (4)$$

Shaw assumes that $n = 1$, and it then follows that at the value of r , where $V(r)/V(o) = 1/2$, $r/D = 1$. Hence

$$D = r(1/2). \quad (5)$$

We shall now comment briefly on some of the assumptions and approximations involved in the derivation of EQUATION 4. First, considering the source and image separately, the relation $v = P/\rho c$, which is valid in the acoustic approximation of weak fields, is not strictly true close in to the charge. It can be shown, however, that the error involved in the peak velocity is of the order of magnitude of $v/2c$, which in all but three of the cases considered does not exceed 2 per cent.

The next approximation is involved in the use of an image-source for the wave reflected at the surface. The combination of source and image does not quite annul the pressure at the surface, there actually remaining a suction ΔP of magnitude $(1/2)\rho V^2$. For a velocity of 200 ft/sec, this amounts to about 18 atmospheres! This suction can be compensated for by adding a system of *positive* image-sources situated above the surface, and treating the latter by the acoustic approximation. The added velocities due to the new system of images (which have both downward and horizontal components) are, however, of the order of magnitude of $\Delta P/\rho c = V^2/2c$. The error in the velocity introduced by neglecting these additional terms is therefore, again, of the order of 2 per cent or less.

As to the assumption that the initial spray velocity is equal to the peak particle-velocity, the following can be said. Let cavitation start when the pressure has been reduced by the reflected wave to a suction of π atmospheres, and let us assume that the pressure varies initially as $e^{-\lambda t}$. It can be shown that when cavitation starts, $\lambda t \sim (\pi + 1)/P$, where P is the peak pressure in the shock wave. The assumption will be justified if $(\pi + 1)/P$ is small, for then the velocities due to both the source and image will not have decreased appreciably from their peak values at the time of rupture. Now P is of the order of magnitude of 200 atmospheres, while π is close to 0 near a solid surface, and is not likely to be much larger in the free water, in view of the probable presence of bubbles in the surface layers. Even adopting Reynolds's estimate of π of about 5 atmospheres, the deviation from the peak velocity, which amounts to about $(\pi + 1)/2P$, is less than 2 per cent.

We may add in passing that when π/P is small, the thickness of the layer above the rupture surface is given by

$$z = \frac{(\pi + 1)cR}{2P\lambda D \left(1 - \frac{nc}{\lambda R}\right)} \quad (6)$$

For a 300-lb. charge of TNT at a depth of 40 feet, $\lambda = 1300 \text{ sec}^{-1}$, and $z(0) < 0.1 \text{ ft}$. The detached layer of water will therefore be of the order of an inch in thickness. This layer will undoubtedly be followed by other layers successively detached.

Modifications of Shaw's method. Alternative methods of determining depth of explosion based on measurements of the initial rate of rise of the center of the dome of spray. In principle, it should be possible to determine the depth either from EQUATION 3

$$V(r) = \frac{2ADW^{n/3}}{\rho c(r^2 + D^2)^{(1+n)/2}} \quad (3)$$

or from EQUATION (4)

$$\frac{V(r)}{V(0)} = \frac{1}{(1 + r^2/D^2)^{(1+n)/2}} \quad (4)$$

If, for example, the initial velocity of the center of the dome $V(o)$ is determined from the photographs, then the depth of explosion D can be inferred from the relation

$$V(o) = \frac{2A}{\rho c} (\Pi^{1/3}/D)^n, \quad (7)$$

where W denotes the weight of charge, and the constants A and n are known for each kind of explosive. Relation (7) depends on a knowledge of the absolute value of the pressure, while relation (4), on which Shaw's method is based, depends only on the relative variation of pressure with distance, or on the shape of the dome of spray.

The difficulties attendant on the application of either of these methods are of a practical nature. The shape or velocity of the dome should be determined from the first few frames of film, but in these initial frames both the height measurements and the velocity measurements are subject to large uncertainties. To begin with, there is an uncertainty in the origin of time (first appearance of dome on the surface), which is of the order of the interval between frames. Secondly, the base-line of the dome is ill-defined in the first few frames, because of the oblique view of the base of the dome which is superimposed on the dome-profile, on account of the elevation of the camera above the water surface. In principle it should be possible to allow for this error when camera height and distance from charge are known, but in some cases the cavitated base extends beyond the full width of the film, and, furthermore, even small errors in measurements of the initial small heights are serious.

The procedure adopted in our analysis was to attach little weight to the measurements in the first third of a second (for a 300-lb. charge at a depth of 30 or 40 feet) and to obtain the initial half-width of the dome by extrapolation from the later frames, in the manner shown in FIGURE 4. Some justification for this procedure may be found in the fact that after about 0.3 seconds the cavitated base has shrunk considerably and the base-line of the dome appears practically straight.

In a similar manner, the initial velocity of the center of the dome $V(o)$ was obtained by extrapolation from a plot of $H(t)/t$, where $H(t)$ denotes the measured height of the center of the dome at time t . It was found that, after about 0.3 seconds, the points align themselves very nearly on a straight line, whose intercept with the axis of ordinates was taken as $V(o)$. The slope of this line determines the deceleration, which was found to be of the order of 2.5g.

The difference between our procedure of measuring the dome shape and Shaw's, is that Shaw measures the heights at a number of points in each of about three frames taken in the first third of a second, while we rely principally on the later frames and measure only the half-widths.

In applying Shaw's method, it is necessary to know the distance scale,

but not the speed of the film. In the analysis by the velocity method, one needs to know both the distance scale and the speed of the film. Now it is a simple matter to determine the film speed by photographing a second-watch on the film (no reliance should be placed on the rated film speed). The determination of the distance scale requires measuring the distance of camera to charge, which is often difficult to carry out. At least, this writer has come across a great many films of good photographic quality in which the distance scale could not be determined, or was uncertain.

In such a case, the depth of explosion can still be determined by combining Shaw's method with the velocity method. The ratio V/D , where D

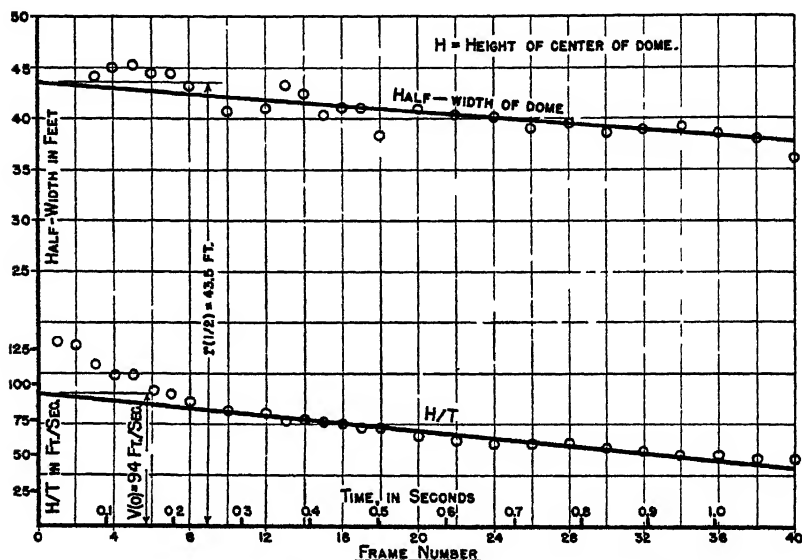


FIGURE 4. Determination of initial half-width and initial rate of rise of dome.

- Film speed determined from a second-watch photographed on film.
- Second-watch fogged or stuck. Film speed assumed equal to average for reel.
- Watch fogged on all shots in reel. Film speed assumed average for all reels

is determined (in terms of an unknown unit of length) from the half-widths, is independent of the distance scale, and the depth can then be inferred from the relation

$$\frac{V(0)}{D} = \frac{2A}{\rho c} \frac{W^{n/3}}{D^{n+1}}. \quad (8)$$

Outline of procedure used in analyzing the dome films. To sum up, the procedure used in analyzing the dome films was as follows:

1. Speed of film was determined from a second-watch photographed on film.
2. The distance scale was computed from the camera-to-charge distance

- (as determined from the cable length paid out), the *measured* focal length of the camera, and the magnification of the film reader used.
- Tracings were made of the dome profiles, and the half-widths $r(1/2)$ then determined, on all frames prior to the emergence of the plumes.
 - The ratio H/t , where H denotes the central height of the dome, and $r(1/2)$ were plotted against t (see FIGURE 4), and from the intercepts of the best-fitting straight lines with the axes of ordinates the initial half-width and velocity $V(0)$ were determined. In passing the straight line through the points, little weight was given to the first ten frames (0.3 seconds) and to the frames immediately preceding the emergence of the plumes.

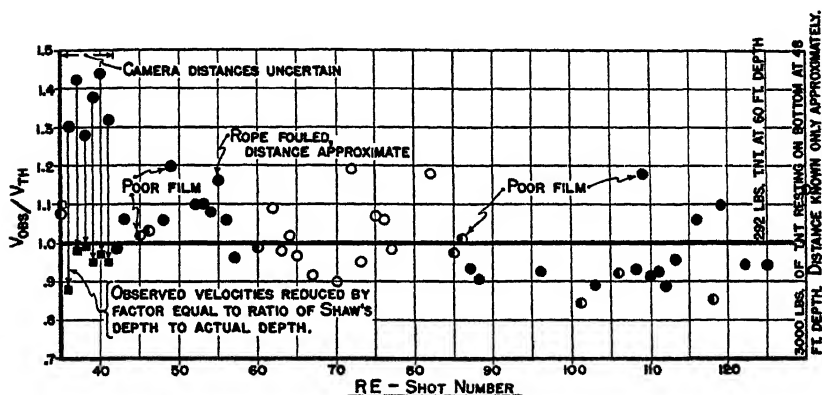


FIGURE 5. Comparison of observed initial rate of rise of center of dome (V_{obs}) with the acoustic particle velocity (V_{th}). V_{th} was taken as $P/\rho c$, where P denotes peak pressure. The latter was taken from the empirically established variation of pressure with distance.

The extrapolated half-width is denoted by $D(S)$. Allowance was then made for the fact that the pressure varies like R^{-n} , $n = 1.15$ to 1.30 (rather than 1.0) depending on the kind of charge. The new depths are denoted by $D'(S)$.

- A depth $D(V)$ was deduced from the equation

$$V = \frac{2A}{\rho c} \frac{W^{n/3}}{D(V)^n}, \quad (7)$$

where W denotes the weight of charge in lbs., and the constants A and n were taken from the empirical formulae for the variation of peak pressure P with distance R ,

$$P(R) = A(W^{1/3}/R)^n. \quad (9)$$

It should be understood that relation 9 is of an empirical nature and is valid in a limited range of $W^{1/3}/R$ only.

- A depth $D(R)$ was computed from the relation

$$\frac{V(0)}{D'(S)} = \frac{2A}{\rho c} \frac{W^{n/3}}{D(R)^{n+1}}. \quad (10)$$

When the distance scale is not known, only $D(R)$ can be determined from the film.

Discussion of results

Observed vs theoretical initial velocities of the spray. The principal result of this investigation, which is new as far as full-sized depth charges are concerned, is a satisfactory agreement between the initial spray velocity and the acoustic particle-velocity based on the empirically determined variation of pressure with charge weight and distance. This is shown in FIGURE 5. In shots 35 through 41 the distances between camera and charge were some-

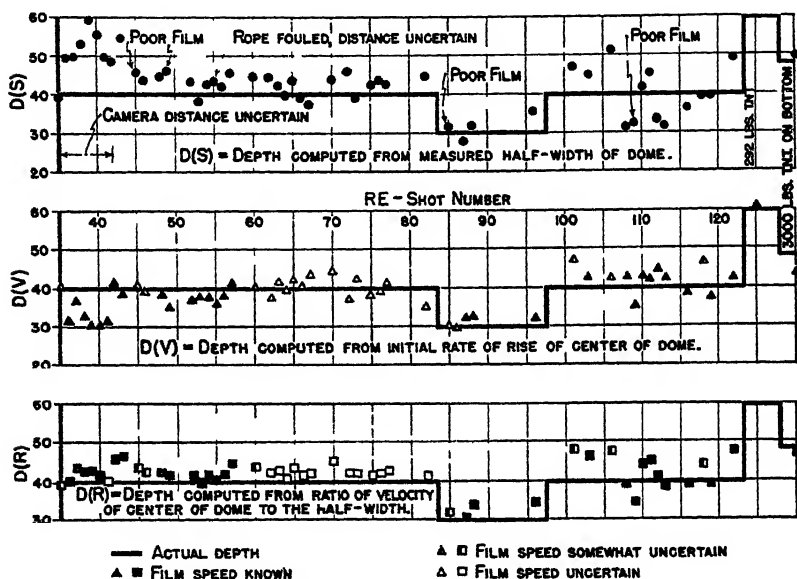


FIGURE 6 Computed versus actual depths.

what uncertain, and it is likely that they were overestimated, because the computed depths $D(S)$ exceeded the actual depths by about the same factor as V_{obs}/V_{th} . When the original values of V_{obs}/V_{th} were reduced by the factors $D(S)/D_{actual}$ the square points were obtained, which are seen to be much closer to unity. The maximum deviation of V_{obs} from V_{th} is about 20 per cent. Some of the extreme deviations occur, however, in shots for which the film speed was not known. In the shots for which both the camera distance and film speed were known, and which were furthermore of good photographic quality, the maximum deviation is 11 per cent.

Some of this deviation may be due to actual variations of pressure with the kind of booster used, as well as to deviations of the wave front from spherical symmetry. In this latter connection, however, it should be pointed out that the group of shots Nos. 75, 76, 119, and 122, in which the

charges were fired with their axes in a horizontal position, is not markedly distinguishable in FIGURE 5 from the other shots in which the charges were fired in a vertical position. All of the charges plotted in FIGURE 5 were suspended in midwater, except the last one of 3000 lbs., which was placed on the bottom. In computing the theoretical spray velocity for this charge the effect of the bottom was not included.

Actual vs. computed depths. The depths deduced from the half-widths $D(S)$, from the velocities $D(V)$, and from the ratio of velocity to depth $D(R)$ are shown in FIGURE 6. Disregarding again shots 35 through 41, for which the distance-scale was uncertain, the maximum deviation of $D(S)$ from actual depth is 35 per cent, that of $D(V)$ 18 per cent, and of $D(R)$ 20 per cent. In the case of $D(R)$, shots 35 to 41 do not appear anomalous because the computation of depth by this method does not depend on the distance-scale.

TABLE 1*

Depth of charge (ft)	Number of shots	Average computed depth by Shaw's method		Average computed depth by velocity method $D(V)$	Average computed depth by (V/D) method $D(R)$	Ratio of measured velocity to $1/3$ velocity	Net charge weight lbs	Depth of water
		$D(S)$	$D'(S)$					
6	1	Distance to camera not known		8.4			18	6
7	1			7.2			100	7
30	4	31.6 \pm 2.0	33.4 \pm 1.8	31.3 \pm 0.9	32.5 \pm 1.0	0.95 \pm 0.03	99-197	80-100
40	46	43.8 \pm 4.2	46.9 \pm 4.4	39.2 \pm 2.7	42.3 \pm 2.7	1.06 \pm 0.09	100-1000	80-100
40	39	42.5 \pm 3.4	45.6 \pm 3.7	40.2 \pm 2.9	42.5 \pm 1.9	1.013 \pm 0.036	100-1000	80-100
48	1	49.1	51.6	43.5	47.0	1.14	3000	48
60	1	61.6	64.7	61.0	63.5	0.94	292	98

*Statistical summary of depths of explosion computed from an analysis of photographs of domes of spray. For an explanation of the different methods used see page 446. In case of the charges fired at 40 feet depth, the 46 shots include all cases analyzed, while in the group of 39 shots, Nos 35 through 41 were excluded, because in these the distance of camera to charge was uncertain.

Besides the shots shown in FIGURES 5 and 6, two others were analyzed for which the distance of camera to charge was not known at all. In such a case, which is apt to occur frequently in practice, the depth can be deduced only from the observed ratio of velocity to the half-width of the dome. In the first case, of a charge of 100 lbs. of TNT resting on the bottom in 7 feet of water, a value of 7.2 was found for $D(R)$. In the second case, of a charge of 18 lbs. of TNT resting on the bottom in 6 feet of water, $D(R)$ turned out to be 8.4 feet. The latter value is excessively high. The only peculiarity of this shot was an unusually large deceleration of the film in the first dozen frames.

A statistical summary of the velocity and depth determination is given in TABLE 1.

Deceleration of the dome. The deceleration of the center of the dome can be estimated from the slope of the velocity curve in FIGURE 4. The average value for 44 shots was found to be more than twice the acceleration of gravity, namely 86 ± 15 ft/sec². With this estimate of the deceleration, one may attempt to speculate on the state of the spray in dome. Two extremes

suggest themselves: (1) the surface of the dome may be impervious to the air, or (2) the spray may consist of individual drops rising through still air with a velocity equal to the observed velocity of the dome. Case 1 can be ruled out immediately because, considering a sheet of water of thickness z feet with a pressure of one atmosphere on top and zero pressure on the bottom, the deceleration would be $(1100/z)$ ft/sec², which would require a thickness of about 20 feet.

In case 2, one can compute the deceleration of a drop when under the influence of gravity and the drag. Taking a typical initial velocity of 80 ft/sec, an observed drag-deceleration of say 50 ft/sec², the diameter of the drop would have to be about 1.5 cm. This is a rather large value, but it is possible that the air partakes in the upward motion, so that the relative motion of the drops is less than the observed motion of the dome, and the drag is correspondingly less.

It may be added that even if the drag-deceleration is proportional to the square of the velocity, the height reached at say 0.8 seconds will be within 10 per cent of the value computed on the assumption of a constant deceleration corresponding to the initial velocity.

Conclusions

The following conclusions have been drawn on the basis of an analysis of photographs of the dome of spray from some 50 statically-fired charges:

1. Estimates of the depth of an underwater explosion from an analysis of the shape of the dome by a modification of Shaw's method can be made with an accuracy of about 10 per cent.*
2. The depth of an underwater explosion can also be determined from the observed initial rate of rise of the center of the dome with an accuracy of about 6 per cent. A technique was developed for determining the initial velocity of the center of the dome, and the measured velocities were found to be within about 7 per cent of the acoustic particle-velocity in the shock wave.
3. When the distance of camera to charge is not known, neither of the above methods can be applied, but the depth of explosion can still be determined from the observed *ratio* of velocity to the half-width of the dome. The depth can be determined by this method with an indicated accuracy of about 5 per cent. A by-product of such a dome analysis is a determination of the distance of camera to charge, which may be of use during testing operations.
4. The satisfactory agreement found between the initial spray velocity and the acoustic particle-velocity in the shock wave offers an independent check (within about 7 per cent) on the absolute pressure measurements made at the Underwater Explosives Research Laboratory, at Woods Hole, Massachusetts.

* This is the range on either side of the correct depths in which are included half the computed depths.

NOTES ON SURFACE WAVES

By MAURICE EWING AND FRANK PRESS

Department of Geology, Columbia University, New York, N. Y.

Introduction. Measurement of the pressure fluctuations at subsurface points, in contrast to direct observation of the sea surface, has recently become the principal method of studying surface waves. The detailed hydrodynamical theory was available long before submarine pressure measurements were made, having been cultivated for its mathematical interest and its bearing on other types of wave propagation. The combination of new methods of observation with an adequate established theory has produced rapid progress during the last few years, and promises much for the future.

The first section of this paper will discuss certain problems in the study of waves by subsurface pressure measurements. The second will present a brief review of some points in classical theory which have been subject to misunderstanding in the recent literature of surface waves. The third outlines a proposed investigation of the swell and microseisms generated by intense storms.

Measurement of Surface Waves by Measurement of Pressure at Subsurface Points

Evolution of pressure recorders for wave studies. Late in 1943, a self-contained pressure recorder, which could sit on bottom and record pressure as a function of time on a scale useful for studying ocean surface waves, was built by Deysher at the Woods Hole Oceanographic Institution, according to a design by Ewing. The main feature of this recorder (see FIGURES 1 and 2) was an external flexible air reservoir which communicated through a slow leak with a rigid case and communicated freely with sensitive bellows within the case. The slow leak amounted to a filter, with time constant about 10 minutes. The bellows were connected with a pen arm which indicated the difference between the instantaneous and the mean pressures. FIGURE 3 shows a typical record obtained with this equipment.

Several instruments of this type were built at the Woods Hole Oceanographic Institution and were used in wave observations by Edmonson, Clarke, and others.

Numerical curves were prepared by Munk for computing wave heights from pressure fluctuations, period, and depth of water according to the classical equations for pressure fluctuation beneath simple harmonic surface waves of small amplitude.

There are many problems in surface waves which could best be solved by a short series of observations at numerous suitable locations, rather than by observations at the few points at which permanent shore-based installations can be made. The important operational advantages for these studies of

an instrument which can be planted on bottom without shore connection, and which can record for several weeks without attention, are extremely great.



FIGURE 1 Self contained pressure recorder

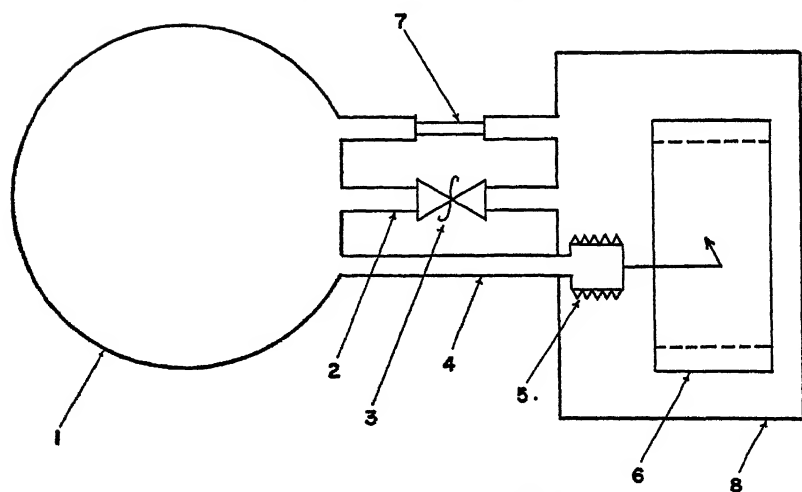


FIGURE 2 * Schematic diagram of self contained pressure recorder

*1 Flexible rubber air reservoir 2 Equalizing tube, for equalizing pressure in reservoir and in watertight box during raising or lowering instrument 3 Safety valves which open when the pressure difference becomes so great that the pressure gauge might be damaged 4 Gauge connection 5 Pressure-sensitive element which measures difference between pressure of air in reservoir and air in watertight box 6 Record chart, driven by clockwork 7 Slow leak 8 Rigid watertight box

It is unfortunate, in our opinion, that the development of this instrument was discontinued, and that it is not available and in use today.

For stations where a long series of observations is desired, the expense and delay of installing a suitable cable and shore-based recorder are often justified. Instruments of this type have been developed by the Woods Hole

Oceanographic Institution, the Naval Ordnance Laboratory, the Admiralty Research Laboratory, Teddington, England, and the University of California.

Relation between spectrum of surface waves and spectrum of subsurface pressure disturbance. Harmonic analyzers have been built at the Admiralty Research Laboratory and at the Woods Hole Oceanographic Institution for the purpose of determining the spectral distribution of the total energy of the surface waves. Several steps are required for this determination. First, the harmonic analysis of the pressure-time record obtained by the subsurface pressure recorder over a representative interval of time (usually taken as 20

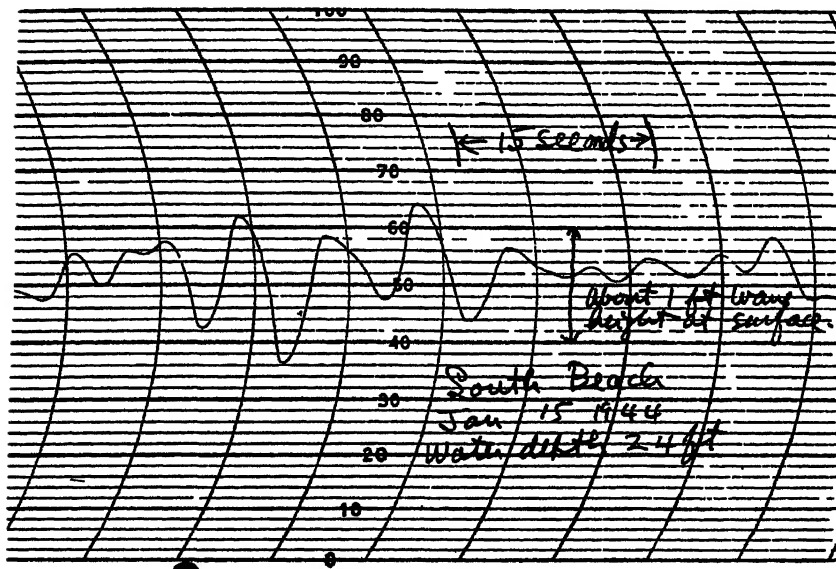


FIGURE 3 Typical record obtained with self-contained pressure recorder

to 30 minutes) must be made. Second, each ordinate on the curve representing the spectral distribution of pressure fluctuation as a function of period, must be multiplied by a factor dependent on period, in order to determine the spectral distribution of the surface displacement.

The first step should be perfectly straightforward, since the theory and design of harmonic analyzers have been well known for many years. However, due to lack of resources and related causes, all of the early analyzers made at the Woods Hole Oceanographic Institution had serious shortcomings. They would not give the proper harmonic components for saw tooth and square waves on test records, and would not properly average the heights of pure sine waves of various amplitudes and periods drawn on test strips. We understand that an improved model will soon be available.

According to available information, the Admiralty Research Laboratory analyzer does not suffer from these difficulties and does give a good approximation to a true Fourier analysis.

The second point, that of multiplying the pressure spectrum by the factor $\cosh kh$ to obtain the spectrum of surface displacements, would have been passed by as apparent from the linearity of the equations, except that some workers in this field have questioned whether this factor should be introduced before or after the frequency analysis of the pressure fluctuation.

Effect of nonrigid bottom on pressure and on velocity. Several independent lines of evidence indicate that the bottom does participate in the wave motion to some extent, whereas all the equations which have been applied in wave studies are based on a perfectly rigid bottom. An effort should be made to appraise the importance of this effect.

Operators of gravity meters placed on bottom in surveys of the continental shelf of the Gulf of Mexico find their instruments disturbed when surface waves are large, no matter how securely they are fastened to bottom or streamlined to decrease resistance to water flow. In his theory explaining the observations of Worzel and Ewing¹ on propagation of compressional waves in shallow water, Pekeris² showed that the participation of the bottom in the wave motion was equally as important as that of the water, and that in all areas investigated the bottom behaved like a liquid. From an excellent experimental investigation of the validity of the classical equation

$$\frac{\Delta H_s}{\Delta H_h} \cosh kh, \quad (1)$$

when $\Delta H_s/\Delta H_h$ is the ratio of the pressure fluctuation near the surface to that at a rigid bottom at depth h , and k is the wave number $2\pi/\lambda$, related to the period τ by the equation

$$\tau^2 = \frac{2\pi\lambda}{g \tanh\left(\frac{2\pi h}{\lambda}\right)} \quad (2)$$

Seiwell³ concluded that wave heights computed from EQUATION 1 were too small by a factor 1.35. The authors investigated his published data to see if the factor 1.35 could be due to yielding of the bottom and found that apparently he had used the deep-water equation

$$\tau^2 = \frac{2\pi\lambda}{g} \quad (3)$$

to relate period and wave number, instead of (2), and that the theoretical curves in his FIGURE 1 were too high as a result. Use of EQUATION 2 would apparently reduce his factor from 1.35 to about 1.18 for waves with period 8 to 9 seconds.*

* Seiwell reported at the conference that the use of EQUATION 3 was intentional, that it reduced the deviation of his equations, and that certain considerations about transformation of deep-water waves entering shallow water justified his choice.

FIGURES 4 and 5 show the corrected curves for rigid bottom drawn through Seiwel's data for Bermuda and Cuttyhunk. In the extreme case of a non-rigid bottom, where the bottom may be treated as a liquid, it has been shown (Lamb⁴) that the pressure fluctuation would fall off with depth h as

$$e^{-kh} = 1/[(1 + \tanh kh) \cosh kh] \quad (4)$$

exactly as in a simple deep liquid. The curves labeled "nonrigid bottom" in FIGURES 4 and 5, are drawn from EQUATIONS 4 and 3, i.e., for the simple liquid case. It will be noted that Seiwel's observed points in these figures

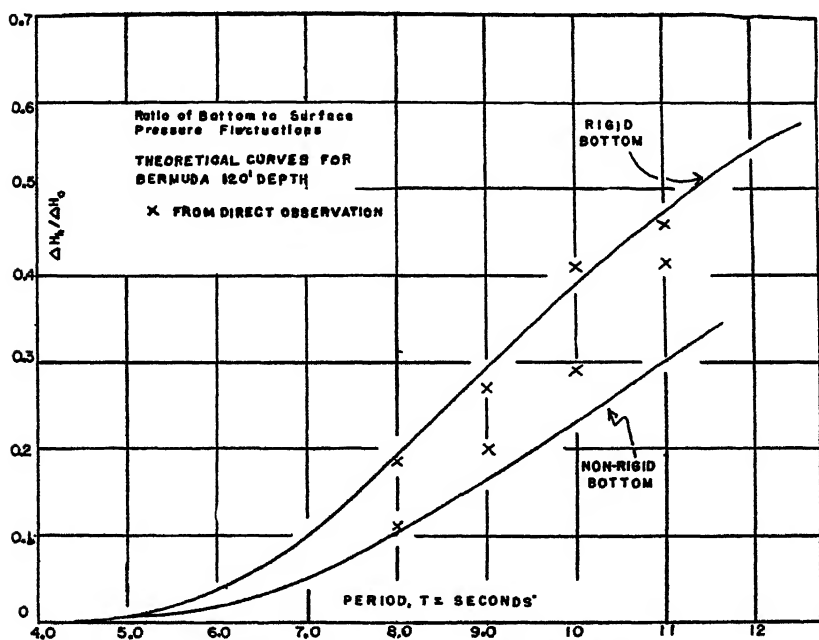


FIGURE 4. Ratio of bottom to surface pressure fluctuations for Bermuda as observed by Seiwel.

lie between the corrected curve for rigid bottoms and the curve for nonrigid bottoms, indicating that a correction for yielding of the bottom produces an effect in the right direction.

Any yielding of the bottom sufficiently great to produce important effects on the pressure will effect, to a like extent, the velocity of propagation. Despite the fact that few measurements of velocity exist, except those incidental to depth measurement in shallow water, where the law $V = \sqrt{gh}$ holds, the absence of reported anomalies in velocity is taken as evidence that yielding of the bottom is at most a second order effect for the types of bottom studied.

In order to obtain quantitative data on yielding of the bottom, it is recommended that determinations of velocity and recordings with seismographs placed beside wave recorders be made. Until this is done, it is desirable

that wave recorders be balanced dynamically, so that they will not respond as seismographs.

Wave measurements from submerged submarines The invention by Vening Meinesz of a pendulum apparatus capable of measuring gravity at sea in a submerged submarine gave, as a by-product, a measurement of the vertical acceleration of the submarine due to wave action, at every gravity station. At the suggestion of Browne⁵, Vening Meinesz⁶ added auxiliary pendulums which also measure the two components of horizontal acceleration due to the waves. The group at Columbia University now making observations of gravity at sea under a contract with the Office of Naval Research, has added a sensitive pressure fluctuation recorder to the gravity equipment, thus tak-

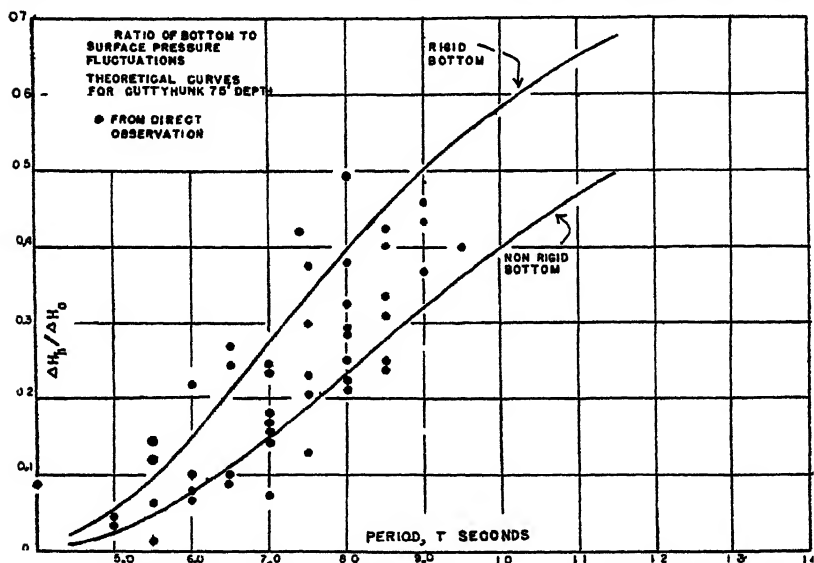


FIGURE 5 Ratio of bottom to surface pressure fluctuation for Cuttyhunk as observed by Serwell

ing the final step necessary to enable a submerged submarine to serve as a very excellent wave-recording station.

If the submarine followed perfectly the particle motion of the water, the acceleration as recorded by the gravity pendulums would give a complete description of the wave motion at any depth where the submarine chose to operate, while the pressure recorder would show no evidence of the presence of waves. If, on the other hand, the submarine was unaffected by the wave motion, the pendulum apparatus would record no wave-induced accelerations, and the pressure recorder would respond like a fixed instrument at the same depth. In all records obtained in the last six months, there is excellent simultaneous data on accelerations and pressure fluctuations from which the amplitude and period of wave motion can be deduced. An example is shown in FIGURE 6.

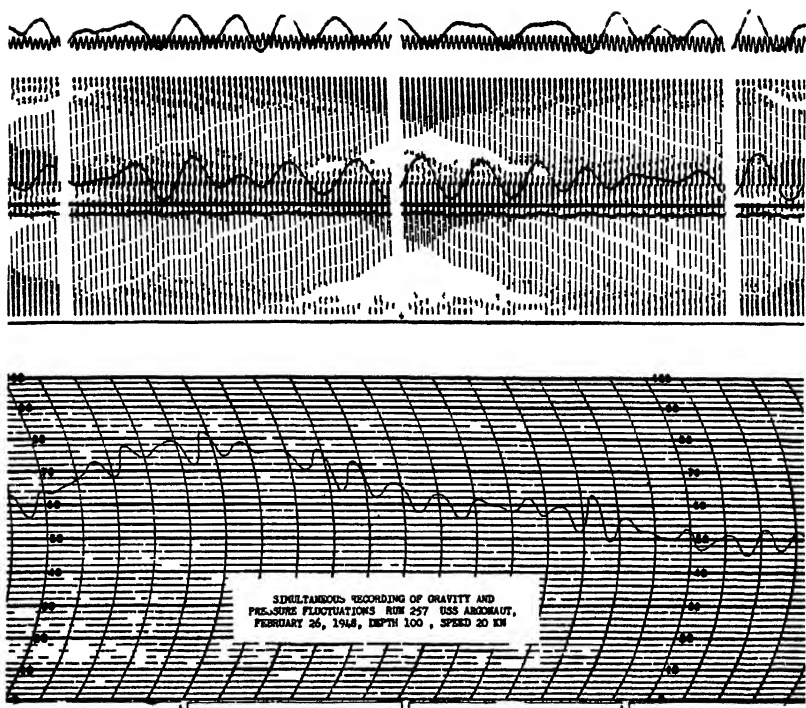


FIGURE 6 Simultaneous records of acceleration and pressure fluctuations obtained during a measurement of gravity at sea in a submarine

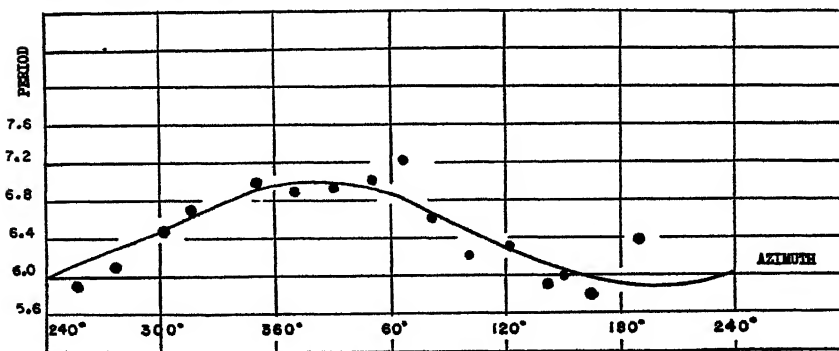


FIGURE 7 Doppler effect on surface wave periods observed on circular run of submarine (USS *Tusk* July 9, 1947, depth 100 feet, speed 15 knots).

In some observations, the course of the submarine has been circular, producing a Doppler effect, from which the velocity of propagation, the period, amplitude, and direction of travel of the waves could be observed, as shown in FIGURE 7.

When more experience has been gained on the extent to which the submarine participates in the wave motion, it will be possible to make excellent wave observations from a submerged submarine using a pressure recorder alone. This will provide an extremely adaptable means for securing data on many difficult problems in wave generation and transmission.

Propagation of Disturbances in Dispersive Media

The splendid revival of activity in the study of surface waves, brought on by the problems of the last war and the resources made available for their solution, has caused a rapid growth in the literature of the subject. Much of the recent literature unfortunately suffers from blemishes of two sorts, misconceptions about some of the fundamental principles, and failure to give sufficiently explicit credit for ideas and results taken from the classical literature. These shortcomings produce a disordered current literature which confuses the beginning student and repels the more seasoned. This situation will ultimately retard progress and should be remedied. The tedious and thankless task of discussing these flaws is undertaken with reluctance, only because they are being quoted and treated as the accepted doctrine on the subject. The principle misconceptions may be grouped into three types:

1. *Elementary descriptions of group velocity*, as typified by the following quotation, from Bigelow and Edmonson⁷, "In the cases, however, of old swells that continue to run on . . . the leading waves tend to die out, chiefly because their energy is expended in setting undisturbed water in motion, but partly because of the resistance of the air which the waves must displace in their advance. The next wave then takes the lead, and this process of replacement continues progressively. Each wave then takes up the energy which was left behind by its predecessor and in turn leaves some of its own energy to be taken up by the next wave." This explanation was evidently not intended to be sufficiently general to cover cases of dispersion, such as flexural waves in a sheet or rod, where group velocity exceeds wave velocity, new waves appear at the front of an advancing group and disappear at the rear, and energy is propagated faster than the individual crests. In addition, the concept of group velocity depends solely on the dispersive nature of the medium and is quite unrelated to energy loss due to the resistance of the air or still water.

2. *Velocity at which energy is carried forward*, as illustrated by the following quotation, from "Breakers and Surf,"⁸ "According to the wave theory only a fraction of the wave energy is carried forward with the wave form, that is, with the velocity C . Let this fraction be called n The numerical value of n approaches $\frac{1}{2}$ in deep water and approaches unity in shallow water." Similar statements appear in Sverdrup and Munk⁹ and in "Wind Waves and Swell."¹⁰ This statement is evidently based on the idea that it makes no difference if one chooses to say that half the energy travels with the phase velocity rather than that all of the energy travels with half the

phase velocity, that is, the group velocity. The weakness of this unorthodox view is apparent when one considers measuring the time required for waves of a given period to travel from an impulsive source to a recording station at a given distance.

3. *Relation between recorded period and period of the underlying spectrum*, as illustrated by the following quotation from Munk¹¹, "The formulation of the problem depends on whether one deals directly with the physical deformations of the transmitting medium, such as we have done, or whether one considers a spectrum of the underlying constant-period wave systems, as in the Cauchy-Poisson analysis. In the first case, one deals with the recorded period, representing the time interval between two successive maxima or minima at a fixed point. In the second case, one deals with component periods which are found by generalizing a given portion of a record into a Fourier integral. In the first case, wave periods increase; in the second case, they remain constant." Elsewhere in the same paper: "It is important to note that the underlying wave system is not in any simple manner discernible in the recorded waves and may indeed be regarded as a mathematical abstraction." These conclusions are reaffirmed in Munk.¹²

These quotations are from a discussion written to explain the well-known increase in period of waves propagated over long distances. It purports to demonstrate that a shift in spectrum from short to long periods occurs solely as a result of propagation in a dispersive medium where the individual waves increase in period and velocity, without any dependence upon selective attenuation. It has long been recognized that, in the orderly sequence of waves set up on the surface of deep water by a localized impulse, the individual waves all increase in length and period. All students have agreed that, in the absence of dissipative losses, the energy associated with each value of the period (in other words, the spectrum) remained constant during transmission, and that the energy was propagated with the group velocity appropriate to the given value of period. The authors consider it valid to represent the forces that generate waves and swell as a summation of localized impulses, and to hold that, in the absence of dissipative forces, there will be no shift in spectrum. As far as the authors can judge, Munk's questionable theories about the increase in period of swell do not enter in any way into his excellent method of tracking storms, which should be judged on its own merits.

Proposed Investigation of Swell and Microseisms

The authors have presented a theory of the generation and propagation of microseisms (Press and Ewing¹³) based on the normal mode propagation of sound in two layers, as developed by Pekeris². The theory indicates the possibility of investigating the geology of the ocean bottom from microseism observations, and suggests several lines of observation to be undertaken.

Since many problems about generation and propagation of swell can be

better solved by numerous simultaneous observations on a single storm than by a few observations on numerous storms, it is proposed that all agencies studying this subject cooperate in an extensive field season for the fall of 1949 or 1950. It is suggested that at least a dozen installations of portable wave and microseism recorders be made on the mainland and on islands to obtain data on all hurricanes passing through the selected area during that season. It is possible that a few key stations, where islands are not available, could be occupied by submarines or surface vessels, if suitable instruments are developed in the meantime.

This investigation would be greatly simplified if self-contained pressure recorders were available by that time.

The inclusion of microseismic studies in this proposal seems natural because of the common origin of the two types of disturbance, and the fact that the same personnel can operate both installations.

Bibliography

1. WORZEL, J. L., & M. EWING. 1948. Long range sound transmission. Sound Transmission in the Ocean. Memoir 27, Geol. Soc. Am., New York.
2. PEKERIS, C. L. 1948. Theory of propagation of explosive sounds in shallow water, in sound transmission in the ocean. Memoir 20 Geol. Soc. Am., New York.
3. SETWELL, H. R. 1947. Investigation of underwater pressure records and simultaneous sea surface patterns. Trans. Am. Geophys. Union 28: 722.
4. LAMB, H. 1932. *Hydrodynamics*. Cambridge University Press.
5. BROWNE, B. C. 1937. The measurement of gravity at sea. Mon. Not. R. A. S., Geophys. Suppl 4: 271.
6. VENING MEINESZ, F. A. 1941. Theory and Practice of Pendulum Observations at Sea, Pt. II. Neth. Geod. Comm. Delft.
7. BIGELOW, H. B., & W. T. EDMONSON. 1947. Wind Waves at Sea—Breakers, and Surf. H. O. Pub. 602, Washington.
8. H. O. Publication. 1944. No. 234 (19).
9. SVERDRUP, H. U., & W. H. MUNK. 1946. Theoretical and empirical relations in forecasting breakers and surf. Trans. Am. Geophys. Union 27: 828.
10. H. O. Misc. 111 (275).
11. MUNK, W. H. 1947. Increase in the period of waves traveling over large distances with applications to tsunamis, swell, and seismic surface waves. Trans. Am. Geophys. Union 28: 198.
12. MUNK, W. H. 1947. Tracking storms by forerunners of swell. J. Meteorol. 4: 45.
13. PRESS, F., & M. EWING. 1948. A theory of microseisms with geological applications. Trans. Am. Geophys. Union 29.

THE PROPAGATION OF SMALL SURFACE DISTURBANCES THROUGH ROTATIONAL FLOW

By PHILIP DUNCAN THOMPSON*

ONR Meteorology Project, The Institute for Advanced Study, Princeton, New Jersey

In the past few years, practical techniques based on the results of wave refraction analysis have been developed for predicting destructive wave energy along shorelines. Concentration of wave energy (as, for instance, the focusing of light) is for the most part due to refraction or variation of phase speed through a nonuniform medium. The argument supporting those methods presupposes that waves which undergo slow changes in form behave like waves of permanent type. As these "quasi-permanent" waves travel shoreward through still water of decreasing depth, their speeds are adjusted to fit that depth according to the classical theory of shallow-water waves. Thus, the configuration of wave fronts and the resulting concentration of wave energy depends entirely, apart from the original properties of the waves, upon the topography of the bottom near shore.

In reality, however, the oceans are never quite at rest, and the assumptions underlying the shallow-water theory, since they are not wholly realized, should be reviewed with regard to wave refraction. The motions of the medium itself (*i.e.*, of tidal currents or streams driven by horizontal variation of density) are reflected in the phase speed. Waves imbedded in a uniform current travel as still-water waves relative to the current itself, but at somewhat different speeds relative to the Earth. If the current is not horizontally uniform, the wave speed varies accordingly and distorts the wave fronts in a refraction pattern. Minute deviations from the still-water phase speed may play a significant part in refraction, simply because they are summed up over finite intervals of time.

Aside from that, all ocean currents are in some degree nonuniform from the surface downward. Indeed, owing to the shear induced by friction between water and bottom, the motion of shallow currents is essentially rotational. The theory of irrotational waves might therefore be expected to fail worst in the circumstances which, as they concern wave refraction, are most critical. For these reasons, it is perhaps worth while to develop a theory of rotational waves, if only to compare the phase speed of waves in a shallow current with that of the corresponding irrotational waves in still water.

The foregoing remarks are intended only to suggest a rather obvious application which might justify a theory of rotational waves, and no attempt will be made to answer the questions of wave refraction. This paper deals with the propagation and, more specially, with the phase speed of small periodic

* The author is indebted to Mr. Gilbert Hunt, ONR Meteorology Project, for valuable suggestions and discussions of this problem, and particularly for pointing out that the depth parameter enters the limiting values of phase speed through the Schwarz inequality.

disturbances traveling through a current whose initial, undisturbed motion is rotational. A method analogous to "fairly shallow water" theory is proposed for finding approximate values of the wave speed in rotational flow. That method represents a solution by a finite series of functions whose coefficients are ascending powers of the wave number, and consequently applies best to long waves.

Several examples of rotational wave motion have been given by Gerstner (Lamb¹), Rankine, and in this monograph by Abdullah, but there appear to be few special instances which admit of exact analysis and very few statements that can be made about the problem as a whole. It is possible, however, to establish the limiting values of the phase speed of rotational waves and to relate them to the speed of the irrotational waves which have been the subject of many earlier and, by this time, classical investigations.

The Boundary Value Problem

In reviewing briefly the theory of irrotational waves, consider motion in the two dimensions normal to a wave front. By cross-differentiating the equations of motion, pressure may be eliminated to reduce the system to two kinematic equations. One of these simply states that no element of fluid suffers dilatation—which is to say, water is incompressible—and the other asserts that the vorticity of any element must remain forever the same. If, in particular, wave motions originate in the state of rest or, more generally, in a state of irrotational motion, the latter implies that those motions are irrotational everywhere and at all times thereafter. This is the most convenient case, for the Cartesian components of potential flow satisfy the Cauchy-Riemann equations. The theory of functions, which deals exclusively with such special variables, is highly developed and accessible, and the boundary value problems that arise in the theory of irrotational waves are not altogether unfamiliar.

On the other hand, if the waves originate in a state of rotational motion, this reduction is no longer available, and we are forced to seek more general solutions of the vorticity equation. It is worth mentioning that, provided the waves are of permanent type, the first integral of the vorticity equation is

$$\nabla^2\psi = F[cx + \psi(x - ct, z)]$$

Here ψ is the stream function; c , phase speed; z , the vertical dimension; x , the horizontal dimension; and t is the elapsed time. There is at least one nontrivial problem—wherein the arbitrary function F is linear—which may be resolved further, but the choice is too limited to be of much physical interest. Furthermore, the free surface condition on wave motions of finite amplitude is nonlinear, and it seems certain that the boundary value problem is more difficult than corresponding problems in the theory of irrotational waves. Rather than to linearize only the boundary conditions, it is reasonable to resort to the method of small perturbations from the very outset.

Before going on to the linear perturbation equations, we shall first consider the undisturbed state of a current, limitless in its horizontal extent and bounded at the bottom by a horizontal plane. The initial, steady state of the current is characterized for simplicity as follows:

- (1) U , the horizontal component of velocity in the x direction, depends only on z , the vertical coordinate.
- (2) Both the horizontal component of velocity normal to the x direction (or y component) and the vertical component are initially zero.
- (3) The initial pressure P depends only on the vertical coordinate, but varies such that the undisturbed current is in hydrostatic equilibrium.
- (4) The undisturbed free surface is horizontal.

It remains to show that this state is physically consistent. First, the velocity of any material element is the same from one moment to the next, which, with the equations of motion, implies that there can be no net force acting to effect changes of velocity. The latter is true by supposition, for the undisturbed current is in hydrostatic equilibrium. Next, the volume filled by any particular mass of water—that is to say, the density—remains unchanged. Finally, because the motion is horizontal, no mass flows through the rigid lower boundary or through the free surface. The pressure at the free surface is constant, of course, and therefore satisfies the dynamical condition of continuity. These arguments are sufficient to show that the undisturbed state is physically possible.

Suppose the current were disturbed by a train of very low waves traveling in the x direction, so that the resulting values of pressure and velocity differ but little from those which characterize the current in its undisturbed state. It is reasonable to believe that the motions at any two points on the same wave front are alike—whence, in other words, the state of the fluid is independent of the y coordinate—and, furthermore, that there is still no component of velocity in the y direction along the wave fronts. In a manner of speaking, the disturbed current is a two-dimensional fluid lying in an (x, z) plane normal to the wave fronts.

If the wave disturbance is sufficiently small, the nonlinear terms in the equations of motion and in the equation of continuity are small beside the linear terms, and may be omitted in the limit. Taking the initial conditions into account at the same time, the linearized equations which govern the motions of the disturbed current may be written in this form.

$$\frac{\partial u}{\partial t} + U \frac{\partial u}{\partial x} + w l'' + \frac{1}{\rho} \frac{\partial p}{\partial x} = 0 \quad (1)$$

$$\frac{\partial w}{\partial t} + U \frac{\partial w}{\partial x} + \frac{1}{\rho} \frac{\partial p}{\partial z} = 0$$

$$\frac{\partial u}{\partial x} + \frac{\partial w}{\partial z} = 0$$

The perturbation variables u , w , p represent infinitesimal deviations from the values of horizontal speed, vertical speed, and pressure in the undisturbed state. The primed quantities have been differentiated with respect to z , the height; ρ is the density of water, which is supposed to be homogeneous. If we now consider only waves of permanent type traveling at constant speed c , local variations with time are explicitly related to variations along the horizontal. After cross-differentiating the equations of motion, u and p may then be eliminated by substitution to obtain the linearized vorticity equation.

$$(U - c)\nabla^2 w = U''w. \quad (2)$$

At this point, we may verify that, if the current were initially at rest, both U and U'' vanish and the differential equations reduce to Laplace's equation, as they do in the Airy theory of irrotational waves.

To state the physical problem completely, we have yet to formulate the boundary conditions. In the first place, any particle immediately adjacent to an impermeable boundary must remain so and, in the present instance, since the bottom is horizontal and rigid, the motions of the bottommost particles are also horizontal. If the origin of the vertical coordinate is chosen to lie at the bottom, the lower boundary condition is

$$[w]_{z=0} = 0. \quad (3)$$

There are two boundary conditions to be satisfied at the free surface. One simply states that the pressure on the lower side of the free surface is equal to the atmospheric pressure on the other side. Since the latter is sensibly constant, we may calculate the pressure at the height of the undisturbed surface. To the present order of approximation, the water contained between the disturbed free surface and the height of the undisturbed free surface is in hydrostatic equilibrium.

$$[P + p]_{z=H} = P_s + \rho g \zeta,$$

where H is the constant depth of the undisturbed stream; P_s is pressure at the free surface; g , the gravitational constant; and ζ is the elevation of the disturbed free surface above the height of the undisturbed surface. Now, $P(x, H, t)$ and P_s are constant and equal by definition, and

$$\left[\frac{1}{\rho} \frac{\partial p}{\partial x} \right]_{z=H} = g \frac{\partial \zeta}{\partial x}.$$

But there is a second condition on the slope of the free surface, for it is always composed of the same sheet of particles. To an observer traveling with the waves, the topmost particles appear to move along the free surface, whence

$$\frac{\partial \zeta}{\partial x} = \left[\frac{w}{U - c} \right]_{z=H}.$$

Finally, inserting this information in (1), the equation of motion in the x direction, the condition equivalent to that at the free surface is

$$\left[(U - c) \frac{\partial w}{\partial z} - \left(U' + \frac{g}{U - c} \right) w \right]_{z=H} = 0. \quad (4)$$

To simplify matters further—although more general solutions may be found by separating variables—we shall confine our attention to periodic waves or, in other words, to solutions of the form

$$w = Z \sin \nu x.$$

The new variable Z depends only on z , the height; ν is the wave number of the periodic oscillation. The circumstances under which w is a solution of the linearized vorticity, EQUATION 2, and satisfies the boundary conditions (3) and (4) are given in the following.

$$Z'' = \left(\nu^2 + \frac{U''}{U - c} \right) Z \quad (5)$$

$$[Z]_{z=0} = 0 \quad (6)$$

$$\left[(U - c) Z' - \left(U' + \frac{g}{U - c} \right) Z \right]_{z=H} = 0. \quad (7)$$

The whole question has thus been reduced to a two-point boundary value problem in ordinary differential equations. The problem is this: given $U(z)$, the distribution of undisturbed flow in a current of depth H , and given the wave number of a small periodic disturbance traveling through it, for what characteristic values of c does $Z(z)$ satisfy EQUATIONS 5, 6, and 7? Those values are the physically admissible speeds at which such waves may travel through the current.

Simple Examples of Rotational Waves and Remarks on Multilayer Methods

Having stated the problem at some length, we shall first find the simplest and probably, with regard to practical application, most useful rotational wave solutions. These examples, which have been mentioned earlier by Haurwitz,² will also serve to illustrate a multilayer method for finding the approximate values of wave speed through a current in which the distribution of undisturbed flow is arbitrary. If the shear in the undisturbed current is constant, the coefficients in the linear EQUATION 5 are constant and the solution can be found by quadratures.

$$Z = A \sinh \nu z + B \cosh \nu z.$$

However, B must be zero, in order that the vertical component of velocity vanish at the bottom, as required by condition (6). The characteristic values of the phase speed are fixed by imposing the free surface condition (7).

$$\left[\nu(U - c) \cosh \nu z - \left(U' + \frac{g}{U - c} \right) \sinh \nu z \right]_{z=H} = 0.$$

In the special case of irrotational flow—*i.e.*, when U' vanishes—the quadratic frequency equation does in fact yield the speed of irrotational waves.

$$c = U \pm \sqrt{\frac{g}{\nu}} \tanh \nu H.$$

Thus, the phase speed of waves in a current whose shear is constant differs from that of irrotational waves, depending on the magnitude of the shear.

Actually, it is rather improbable that the shear in a shallow stream is constant throughout. What is more likely, friction-induced shear is great next to the bottom and small near the surface. Suppose that the shear is constant below some definite level and vanishes above, suffering a sharp discontinuity at the level itself. To resolve the two-layer problem, however, we must first establish rules for integrating across the point where U'' is undefined—that is, for piecing separate solutions together at that point. The necessary information is contained in the statements that pressure and vertical speed are continuous across the shear discontinuity. Since pressure is continuous, the integrated pressure difference between two points on the shear discontinuity must be the same, whether the path of integration lies wholly on one side of the discontinuity, or whether it lies wholly on the other. This, with EQUATION 1, implies that

$$[(U - c)(Z'_2 - Z'_1) - Z_1(U'_2 - U'_1)]_{z=D} = 0 \quad (8)$$

$$[Z_2 - Z_1]_{z=D} = 0. \quad (9)$$

The subscripts 1 and 2 refer to conditions below and above the discontinuity, which is at height D above the bottom. In the present case U'_2 vanishes.

The differential EQUATION 5 may be integrated within the interior of each of the layers above and below the surface of discontinuity.

$$Z_1 = A_1 \sinh \nu z + B_1 \cosh \nu z$$

$$Z_2 = A_2 \sinh \nu z + B_2 \cosh \nu z.$$

Again, B_1 vanishes. The constants A_2 and B_2 are fixed by conditions (8) and (9). The value of A_1 is immaterial. Since all the conditions are linear homogeneous, A_1 occurs in the characteristic equation only as a non-zero factor. Finally, we may find the characteristic values of phase speed by imposing the free-surface condition (7) on Z_2 .

$$\left[\nu(U - c)^2 \cosh \nu z - U'_1(U - c) \sinh \nu D \cosh \nu(z - D) - g \sinh \nu z + \frac{gU'_1}{\nu(U - c)} \sinh \nu D \sinh \nu(z - D) \right]_{z=H} = 0.$$

This equation, which is cubic in $(U - c)$, can be solved explicitly only after

considerable labor. However, something about the nature of the roots may be discovered by inspecting its behaviour as U'_1 becomes small. First, if U'_1 is zero, the characteristic values of the phase speed are those of irrotational waves traveling upstream and downstream. Also, as U'_1 becomes small, the term of zero degree vanishes and the remaining root approaches zero. For reasonably small values of U'_1 , the conjugate roots are opposite in sign, whence we may infer the sign of the third root. If, for instance, U'_1 is positive, the product of the roots must be negative, and the third root is positive. The reverse is true, of course, if U'_1 is negative. The results of this estimate are twofold: first, a weak shear discontinuity introduces an additional characteristic value in the neighborhood of the current speed at that discontinuity, and second, the distribution of vertical speed corresponding to that value is extremely sensitive to small variations of the characteristic value.

Since the methods for handling two-layer problems can be extended to apply to any number of layers, it is natural to represent an arbitrary profile of current flow by a sequence of short, connected straight lines. This measure is dictated by reason as well as convenience, for our observational knowledge of the distribution of flow is of precisely this kind. In brief, the method is this. Imagine that the undisturbed current is composed of a number n of thin horizontal layers, separated by $(n - 1)$ surfaces of shear discontinuity. In each of those layers the shear is constant and the solution, except for the values of two arbitrary constants, is completely known; thus there are in all $2n$ constants to be fixed. One of these is determined by the kinematic boundary condition at the bottom, and $(2n - 2)$ constants are fixed by $(2n - 2)$ conditions, similar to (8) and (9), at the $(n - 1)$ surfaces of discontinuity. Finally, since it occurs in the characteristic equation only as a nonzero factor, the one remaining constant is clearly unimportant; consequently the free-surface condition really contains no arbitrary constants and determines the phase speeds admissible under given conditions.

Now, the degree of the characteristic equation increases with the number of layers, and we are faced with the problem of detecting which of the superficially admissible values of phase speed are real and which are spurious values introduced by fictitious and necessarily weak shear discontinuities. In the first place, if the shear is really continuous, the characteristic values in the range of the current speed are nonexistent—for, if they were not, EQUATION 5 would imply that the vertical speed is somewhere discontinuous, which is impossible. Furthermore, because the artificial discontinuities are weak, the distributions of vertical speed corresponding to the spurious values are most sensitive to small variations in characteristic value. This property enables us to filter out, by increasing the order of approximation, the characteristic roots that have been introduced by the method itself.

Long Rotational Waves

The boundary value problem presented in EQUATIONS 5, 6, and 7 may be stated more concisely by introducing a nondimensional variable

$$\phi \equiv \frac{Z}{U - c}.$$

The governing differential equation and the boundary conditions then reduce to

$$(K\phi)' = \nu^2 K\phi \quad (10)$$

$$[K\phi' - g\phi]_{z=H} = 0 \quad (11)$$

$$[\phi]_{z=0} = 0 \quad (12)$$

where, for convenience, $K \equiv (U - c)^2$. The form of the differential EQUATION 10 suggests that its solution might be easily expressed as a summation of functions whose coefficients are ascending powers of the wave number. If the wave number is small—i.e., if the wave length is great—such a series is rapidly convergent. We shall suppose that the wave number is very small, small enough that the series solution may be terminated after the first two terms.

$$\phi = \phi_0 + \nu^2 \phi_1$$

Now, since ϕ is a solution of (10)

$$(K\phi_0)' = 0$$

$$(K\phi_1)' = K\phi_0.$$

However, the wave number is small by supposition, so that ϕ is almost ϕ_0 and

$$(K\phi_1)' = K\phi_0.$$

Thus, ϕ_0 and ϕ_1 may be found by successive quadratures. Subject to condition (12),

$$\phi_0 = \int_0^z \frac{dz_1}{K}$$

$$\phi_1 = \int_0^z \frac{dz_1}{K} \int_0^{z_1} K dz_2 \int_0^{z_2} \frac{dz_3}{K}.$$

Finally, the long wave solutions of (10) are

$$\phi = \int_0^z \frac{dz_1}{K} + \nu^2 \int_0^z \frac{dz_1}{K} \int_0^{z_1} K dz_2 \int_0^{z_2} \frac{dz_3}{K}.$$

As before, the characteristic values of phase speed are fixed by the free surface condition (11).

$$1 + \nu^2 \int_0^H K dz \int_0^z \frac{dz_1}{K} - g \int_0^H \frac{dz}{K} - g\nu^2 \int_0^H \frac{dz}{K} \int_0^z K dz_1 \int_0^{z_1} \frac{dz_2}{K} = 0. \quad (13)$$

To illustrate the way in which phase speed is determined by EQUATION 13, we shall consider the special case of irrotational waves. In that event the shear is zero and U is constant; then K is also constant, and the appraisals may be made in finite terms.

$$1 + \nu^2 \int_0^H dz \int_0^z dz_1 - \frac{g}{K} \int_0^H dz - \frac{g\nu^2}{K} \int_0^H dz \int_0^z dz_1 \int_0^{z_1} dz_2 = 0.$$

Carrying out the indicated integrations and replacing K with its defined value

$$(U - c)^2 = gH \left[\frac{1 + \frac{(\nu H)^2}{6}}{1 + \frac{(\nu H)^2}{2}} \right].$$

If νH is small—that is to say, if the wave length is large compared with depth—then

$$\frac{1 + \frac{(\nu H)^2}{6}}{1 + \frac{(\nu H)^2}{2}} \approx 1 - \frac{(\nu H)^2}{3}$$

$$(U - c)^2 = \frac{g}{\nu} \left[\nu H - \frac{(\nu H)^3}{3} \right].$$

The bracketed factor on the right is the sum of the first two terms in the series expansion for the hyperbolic tangent of νH , which converges rapidly when νH is small.

$$c \approx U \pm \sqrt{\frac{g}{\nu} \tanh \nu H}.$$

This result is in accord with the classical theory of irrotational waves.

If, on the other hand, the distribution of flow in the undisturbed current is arbitrary, we must resort to iteration methods for finding the characteristic phase speed. An ordered sequence of approximations c_0, c_1, \dots, c_n may be calculated by Newton's method. The left hand side of (13), $f(c)$, is first differentiated with respect to c , omitting from the derivative those terms of which the wave number is a factor, but retaining them elsewhere. This is justified, for, according to our hypothesis, the wave number is small. Next, c_{i+1} , the $(i + 1)$ st order approximation to a zero of $f(c)$, is built up from $f(c_i)$ and $\frac{df}{dc}(c_i)$ by Newton's method. Whether or not this method is successful depends largely

on the quality of the first approximation, about which the method itself tells nothing. However, since the wave motions are irrotational in the first approximation, we might hazard a reasonable guess by choosing the speed of irrotational shallow-water waves.

Limiting Values of Phase Speed

The method outlined in the foregoing section applies, strictly speaking, only to periodic disturbances whose wave length is fairly large compared with the depth of the current. For aesthetic reasons, as well as to reassure ourselves that rotational waves cannot travel at speeds wholly unexpected from approximate results, we should prefer more general statements about the phase speed. It is possible, in spite of our inability to integrate the differential equation completely, to fix the limiting values of phase speed under rather general conditions. To show this, we shall first multiply both sides of EQUATION 10 by ϕ and integrate from the bottom to the height of the undisturbed free surface.

$$\int_0^H \phi(K\phi')' dz = \int_0^H v^2 K \phi^2 dz$$

Because the integrand of the integral on the right hand side is never less than zero, the integral itself is never less than zero, whence

$$\int_0^H \phi(K\phi')' dz \geq 0.$$

Integrating by parts

$$[\phi K \phi']_{z=0}^{z=H} - \int_0^H K (\phi')^2 dz \geq 0$$

which, subject to conditions (11) and (12), becomes

$$[g\phi^2]_{z=0}^{z=H} - \int_0^H K (\phi')^2 dz \geq 0.$$

Also, by reason of (12),

$$g \left[\int_0^H \phi' dz \right]^2 - \int_0^H K (\phi')^2 dz \geq 0.$$

Now, in virtue of the Schwarz inequality,

$$H \int_0^H (\phi')^2 dz \geq \left[\int_0^H \phi' dz \right]^2$$

and finally,

$$\int_0^H (gH - K) (\phi')^2 dz \geq 0. \quad (14)$$

For the sake of argument, let us now suppose that c is greater than $(U_{\max} + \sqrt{gH})$. If so, then K , which is $(\bar{U} - \bar{U}_{\max} - \sqrt{gH} - \epsilon^2)$, is always greater than gH . But in that case, the integrand of (14) is always less than zero, which violates the inequality. The same contradiction arises if c is chosen less than $(U_{\min} - \sqrt{gH})$. The physically possible values of phase speed must therefore lie within the range defined by the following inequalities.

$$U_{\min} - \sqrt{gH} \leq c \leq U_{\max} + \sqrt{gH}$$

One already well-known result follows immediately from this inequality: the limiting speed of low irrotational waves is that of very long "shallow-water" waves.

By imposing weak restrictions on the distribution of undisturbed flow, the maximum speed of waves traveling downstream may be related to the speed of irrotational deep-water waves. We shall suppose that U is a monotonic and, which is likely in reality, nondecreasing function of height. After multiplying both sides of EQUATION 10 by the factor $2K\phi'$, we shall integrate from the bottom to the height of the undisturbed free surface, as before.

$$\int_0^H 2K\phi'(K\phi')' dz = \int_0^H \nu^2 K^2 2\phi\phi' dz.$$

Integrating by parts,

$$[(K\phi')^2]_{z=0}^{z=H} = [\nu^2 K^2 \phi^2]_{z=0}^{z=H} - \nu^2 \int_0^H \phi^2 (K^2)' dz$$

which, subject to boundary conditions (11) and (12), reduces to

$$[\phi^2 (g^2 - \nu^2 K^2)]_{z=0}^{z=H} = [(K\phi')^2]_{z=0}^{z=H} - \nu^2 \int_0^H \phi^2 (K^2)' dz.$$

Finally, replacing K in the integral on the right hand side with its defined value,

$$[\phi^2 (g^2 - \nu^2 K^2)]_{z=0}^{z=H} = [(K\phi')^2]_{z=0}^{z=H} - 4\nu^2 \int_0^H \phi^2 (U - c)^3 U'' dz. \quad (15)$$

Now, let us examine the consequences of supposing that c is greater than $(U_{\max} + \sqrt{g/\nu})$. In the first place, $(U - c)^3$ is always less than zero, for $(U - U_{\max} - \sqrt{g/\nu})$ is negative. By supposition, U' is never less than zero. Taken together, the latter imply that the integrand of the integral on the right-hand side of (15) is never positive, and that the right-hand side itself is positive or zero. In other words, if

$$c = U_{\max} + \sqrt{g/\nu} + \epsilon^2,$$

then

$$[\phi^2(g^2 - \nu^2 K^2)]_{z=0} \geq 0;$$

whence

$$\left(\frac{g}{\nu}\right)^2 - [U - c]^4]_{z=H} \geq 0.$$

This conclusion is obviously impossible, since, according to the very same supposition,

$$\left(\frac{g}{\nu}\right)^2 - \left[\left(U - U_{\max} - \sqrt{\frac{g}{\nu} - \epsilon^2} \right)^4 \right]_{z=H} < 0.$$

Because the assumption tentatively adopted above leads to contradictory propositions, its negation is surely true.

$$c \leq U_{\max} + \sqrt{g/\nu}.$$

These results, though they tell little about rotational wave propagation, are exact and of more than casual interest, as they concern numerical solution of the boundary value problem. If it is known that the characteristic values of phase speed must lie within a calculable finite range, there is certainly no need to look further for values which can be tested only at great expense.

References

1. LAMB, H. 1932. *Hydrodynamics* 6th Ed. Cambridge University Press. London (Dover Publications. New York. 1945): 421-423.
2. HAURWITZ, B. 1931. Zur Theorie der Wellenbewegungen in Luft und Wasser. Veröffentlichungen des Geophysikalischen Instituts der Universität Leipzig Zweite Serie. 5(1): 63.

RECENT STUDIES OF WAVES AND SWELL

By G. E. R. DEACON*

Admiralty Research Laboratory, Teddington, England

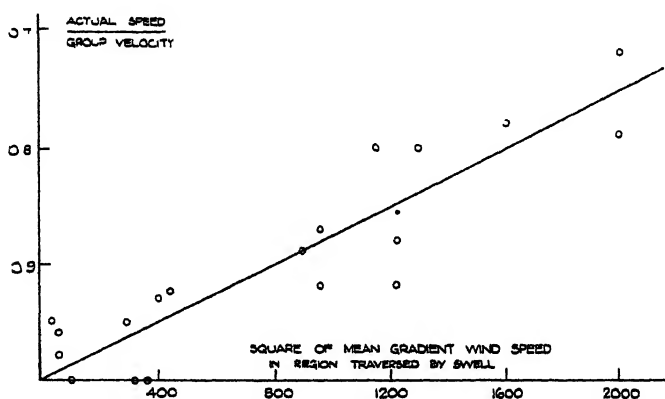
Generation and propagation of waves. The examination of wave spectra, obtained by submitting wave records to frequency analysis, shows that the mixture of waves generated in a storm comprises waves of all lengths up to a maximum which depends on the greatest wind strength. By comparing the measured times at which the different wave lengths begin and cease to arrive at a recording station, with the times (estimated from meteorological charts) at which they began and ceased to be generated in the distant storm area, it has been shown that the component wave trains behave independently of each other, and each train advances across the ocean at a speed which is within 5 per cent of the theoretical group velocity appropriate to its period. The short waves are, therefore, overtaken and outdistanced by the longer waves, and the separation between them increases with distance. In an individual spectrum at a very distant recording station the swell covers only a narrow range of periods, and, after allowance has been made for the effect of the tidal streams in the coastal region on the apparent wave period as registered by a stationary instrument, it can be shown that the upper and lower limits of the narrow wave band decrease slowly and continuously with time. At a recording station closer to the generating area the wave band is wider, and it broadens rapidly toward shorter periods.

The trend toward shorter periods can be used to estimate the distance of the generating area from the recording station, though a precise answer can be expected only when the generating area is small. For example, a band of swell of 18 to 20 seconds period appearing on spectra from the coast of Cornwall can be shown to have been generated in a hurricane off the coast of Florida; and bands of swell characterized by their very slow trend towards shorter periods, shown to have their origin in the stormy latitudes of the South Atlantic Ocean. The techniques used, and a theoretical basis to the work, are described in a paper by Barber and Ursell¹.

Effect of following winds. The deduction that waves of a particular period advance across the ocean at a speed which approximates very closely to the theoretical group velocity is based on the timing of swell from small, intense, storms at a great distance. Since such wave bands appear most plainly in the available spectra from the coast of Cornwall when no waves are arriving simultaneously from the eastern half of the Atlantic Ocean, there has been a tendency to select conditions in which the swell travels through relatively calm water. An attempt is now being made to time swell which travels through the eastern half of the ocean when this is disturbed by strong winds.

* The work described in this summary is the result of close collaboration between members of the Oceanographical Research Group at the Admiralty Research Laboratory, Teddington, and they are indebted to the Admiralty for permission to publish it.

Up to the present, there are sufficient data to study the effect of following winds only, and although these would be least expected to retard the swell, they appear to slow it down by as much as 20 per cent if the gradient wind speed between the generating area and the coast averages as much as 40 knots. There seems to be an approximately linear relationship between the reduction in the rate of travel of the swell and the square of the wind velocity, which gives some indication that the retardation is due to increased turbulence in the water through which the swell travels. A graph showing the ratio of actual to theoretical velocities plotted against the square of the gradient wind speed is reproduced in FIGURE 1, but until further data have been examined and a suitable explanation has been found the result is viewed with some suspicion.



publication, N. F. Barber shows that the average length of the waves in a group expands or contracts at the same rate as the general surface of the water in which the group is moving, the changes in wave length being produced by the expansion and contraction of the water surface. It follows that the change in apparent period will be proportionately small when the waves complete their passage through the tidal area in a small fraction of a tidal cycle, and the correction which must be made to the measured periods to obtain the true periods will be smaller where the tidal streams are weak and narrow.

Up to the present, only narrow bands of swell from distant storms have been examined for variations in period due to tidal streams. Similar variations probably take place in all waves and swell crossing the tidal area, but they are not likely to be detected so easily nor to possess the same significance

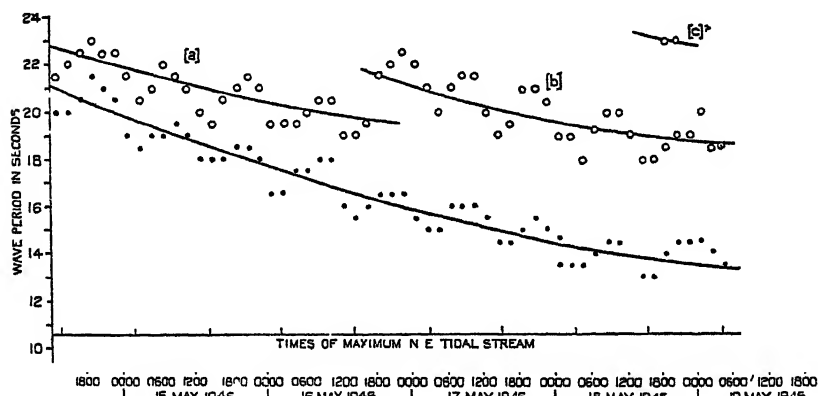


FIGURE 2. The maximum and minimum periods limiting a narrow frequency band due to swell from a storm area in the South Atlantic Ocean (after BARBER and URSSELL).

in swell which has just left the generating area, or in waves of local origin, because of other variations due to the changing character of the storm.

Because of the long travel time and separation of the component wave trains of swell arriving at a very distant recording station, it can be expected that the measured height should be subject to only slow and continuous variations. The measurements from the coast of Cornwall show variations which appear too large and too rapid to be attributed to a distant cause, such as the variations in the storm area. Two examples, showing the measured height of swell from distant storms plotted against the time of arrival, are reproduced in FIGURE 3. They give some indication of a 6-hour periodicity, and a numerical test shows a component at 6 hours and a smaller one at 12 hours. Some fluctuation of semidiurnal period might be expected because of different refraction of the waves in shallow water at high and low tides, and also because inadequate allowance may have been made for the varying

depth of water over the recording instrument, but no satisfactory explanation has been found for the apparent 6-hour period. There is some indication in the graphs that the greatest reduction of swell-height occurs at the time of maximum tidal stream at the recording station, and it is possible that the swell may lose energy in passing through an area of turbulent streams. The question is of some importance in connection with the use of swell recordings to study the generation of waves in distant storms, and a method is being devised to make continuous observations of the height of narrow bands of swell.

Wave prediction. With a view to obtaining information likely to be of direct use in improving the methods of predicting waves and swell, attempts are being made to predict the wave spectrum on the coast of Cornwall.

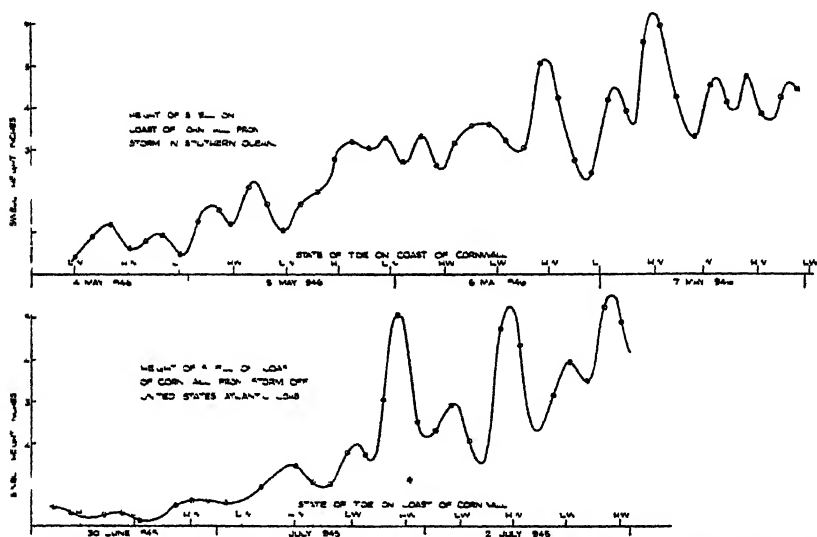


FIGURE 3. The variation in the height of low swell reaching the coast of Cornwall from a storm off the United States Atlantic coast, and from a storm in the South Atlantic Ocean

The width of the spectrum when waves are being generated by a rising wind suggests that the energy begins to be distributed over a range of wave lengths as soon as the waves are formed, some being communicated to the longer wave lengths before the shorter ones are fully energized. Each spectrum has an optimum band in which the waves are highest. Waves shorter than this optimum period are lower, presumably because they have a smaller capacity for absorbing energy without becoming unstable; and longer waves are lower, probably because their speed, which is nearly equal to that of the wind, allows them less opportunity of absorbing energy. A comparison of the wave characteristics and wind speeds shows that the period, in seconds, of the longest waves occurring in the spectrum has a numerical value approximately one third of the gradient wind speed, in

knots, in the strongest part of the storm. The period of the highest waves is approximately 25 per cent less than the period of the longest waves. These empirical rules appear fairly reliable for North Atlantic storms, but they may have to be modified for very small generating areas. Reasonable assumptions have been made about the effect of the wind on the wave characteristics, and of the capacity of individual wave lengths for absorbing energy as the wind strength increases, and for retaining it as they travel through and beyond the storm area. The resulting conclusions are being tested against the empirical rules and recorded wave spectra. It is hoped that a preliminary report will soon be ready for publication.

Practical applications of swell recording. The results obtained up to the present in the United States and Great Britain make it possible to use swell recordings from a coastal station to estimate the strength of wind in the generating area, and its distance from the recording station. This indirect information is not likely to prove very useful in the North Atlantic Ocean, where there are plenty of meteorological observations, but it may be useful in the southern hemisphere, where many coasts are exposed to oceans from which little meteorological information can be expected. The existing techniques can be improved, particularly by the development of an instrument to record the wave spectrum at frequent intervals and of a method for measuring the direction of long swell while it is obscured by a pattern of shorter waves but, without any further development, they could be put to good use on coasts where communications and the transshipment of cargo are hindered on "swell days," or at flying-boat stations from which the boats must take off before the swell grows too high. The installation of swell-recording equipment at such places would soon pay for itself, and make a useful contribution to local knowledge, and to the general study of waves and swell.

The reasonable certainty of detecting an oncoming swell as much as twelve hours before it can be seen may prove to be useful as a warning of the approach of cyclones in places which have to rely to a large extent on local indications. In such an application, swell recording may not be able to compete with the microseismic direction-finding methods used by the United States Navy in the "Hurricane Project," but for the time being, it seems very important that the waves and microseisms should be studied together.

Relations between sea waves and microseisms. The cooperation of the Superintendent of Kew Observatory has made it possible for wave records from Perranporth on the coast of Cornwall to be compared with simultaneous recordings of microseisms at Kew, a western suburb of London. Measurement of the records appeared to show such close agreement between the variations in wave and microseismic activity as to suggest that the microseisms are due mainly to wave activity in the coastal region, and not, to any appreciable extent, to the passage of cyclonic depressions over deep water

west of the British Isles. This result appeared contradictory to the findings of the Hurricane Project, which showed that in Puerto Rico, Cuba, and Florida the effect of wave activity in the coastal region was negligible compared with that of deep barometric depressions over the ocean.

New light was thrown on the subject when it was found that the dominant wave period at Perranporth was double the dominant microseism period at Kew, and when it was observed that Bernard² had found the same relation between wave and microseism periods at Casablanca. Bernard suggested that microseisms may be produced in places where the interference between wave trains gives rise to standing oscillations, and Longuet-Higgins and Ursell, in a paper which is being prepared for publication, have pointed out that Miche³ has shown that the average pressure on the sea bottom below a standing wave varies sinusoidally with twice the frequency of the wave, and with an amplitude which does not decrease to zero with increasing depth. Using the work of Miche, and expressing the problem more precisely than Bernard, Longuet-Higgins is preparing a new theory of the generation of microseisms in which the essential requirement is that there should be interference between waves of the same period traveling in opposite directions, such as may result from the reflection of swell from a coast or the interference of waves generated in opposite quadrants of a cyclonic depression. Where such interference takes place, the mean pressure on the sea bottom over the whole area varies with twice the frequency of each wave train, the amplitude of the variations being proportional to the product of the amplitudes of the separate wave trains. Such a theory affords a satisfactory explanation of how pressure variations, of the same period as the microseisms, can be produced over relatively large areas of the sea bottom, and it should stimulate further work on the relations between sea waves and microseisms.

With such good reasons for believing the period of the microseisms to be so closely related to that of the waves, the problem of distinguishing the effect of a cyclonic depression passing over deep water west of the British Isles on the microseisms at Kew, in the presence of a stronger effect due to waves in the coastal region, resembles the problem of distinguishing swell from a distant storm in the presence of waves of local origin. If there is an appreciable difference between the dominant wave periods in the two areas, there should be a corresponding difference between the periods of the microseisms, and it should be possible to distinguish between microseismic waves from the two sources by submitting a seismograph record to frequency analysis.

The small amplitude of the microseisms, as recorded by one of the established seismographs, makes it difficult to obtain a frequency analysis; but a method, using an automatic curve-follower⁴ to reproduce the seismograph record in a form suitable for analysis, has been developed. An example, showing simultaneous wave spectra from Perranporth and microseism

spectra from Kew is reproduced in FIGURE 4. Between 1200 hrs., 18 October, and 0500 hrs., 19 October, swell with a mean period of 10 to 12 seconds dominates the wave spectra and microseisms with a mean period of 6 to 7 seconds dominate the microseism spectra. After 0500 hrs., 19 October, there is a second band of microseisms with a mean period of 6 to 7 seconds, and the absence of a corresponding increase in activity in the wave spectra shows

COMPARISON OF WAVE SPECTRA AT PERRANPORTH WITH THE SPECTRA OF MICROSEISMS
RECORDED AT THE SAME TIME AT KEW

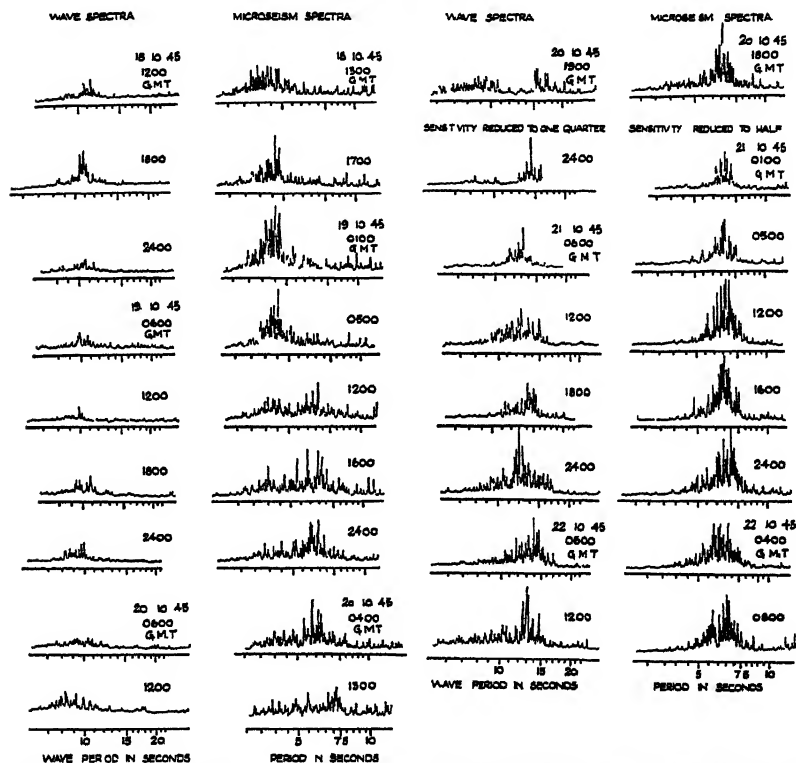


FIGURE 4. Comparison of wave spectra from the coast of Cornwall with microseism spectra from Kew.

that the new band of microseisms cannot be attributed to new swell in the coastal region. It can, however, be attributed to wave activity in a depression south of Greenland during the night of 18 to 19 October, since there was no other wind area capable of generating waves of 12 to 14 seconds period. The swell from the area south of Greenland began to arrive at Perranporth between 1200 hrs. and 1800 hrs., 20 October, and its arrival in the coastal region was the signal for a further increase in microseismic activity of 7 seconds period, first apparent in the spectrum for 1300 hrs. The period of the highest waves arriving from the storm south of Greenland is approxi-

mately twice that of the microseisms. This example is not regarded as unmistakable evidence that microseisms are generated in deep water, since those attributed directly to the storm area may have been generated in coastal waters near Greenland; but further work which is being done on the same lines appears more conclusive. It is perhaps worth noting that increases in the microseisms at Kew generally precede increases in wave height at Perranporth by about 6 hours, as though the microseisms were generated farther west than Perranporth. It is hoped that further details of this work by Longuet-Higgins, Darbyshire, and others were communicated to the assembly of the International Union of Geodesy and Geophysics in Oslo. It is believed that sufficient work has been done to show that the precise and comprehensive data necessary for a detailed study of the generation of microseisms can only be obtained if microseisms and waves are studied together, and to show that microseisms may prove to be useful in predicting swell.

References

1. BARBER, N. F., & F. URSELL. 1948. *Phil. Trans. Roy. Soc.* **240**: 527.
2. BERNARD, P. 1941. *Bull. Inst. Oceanographique, Monaco*, **38**: 800.
3. MICHE, M. 1944. *Ann. Ponts & Chaussées* **114**: 25.
4. TUCKER, M. J., & G. COLLINS. 1947. *Electronic Engineering* **19**: 398.

SEA SURFACE ROUGHNESS MEASUREMENTS IN THEORY AND PRACTICE

By H. R. SEIWELL*

Woods Hole Oceanographic Institution, Woods Hole, Mass.

Introduction. Although the physical characteristics of the sea surface have influenced the economy and livelihood of a large percentage of mankind, it is a subject which has rarely been investigated systematically. It is perhaps no exaggeration to state that catastrophic conditions of the sea have claimed as many lives and destroyed as much property as any other geophysical phenomenon. However, regardless of its hazardous history, the sea surface is often regarded with a sort of mysticism as a natural phenomenon to be gambled with, rather than approached in an orderly systematic fashion. To some extent, a disinterested attitude toward marine scientific investigations has arisen from an absence of ocean-mindedness and from a false security of this country's isolation by the world's oceans. Thus, industries frequently proceed from the standpoint of overcoming the natural forces of the ocean, rather than approaching them in a harmonious manner.

The basic need in a general wave research program is systematic and accurate observations on the state of the sea surface. The present status of wave research is somewhat analogous to that of oceanic circulation a half century ago. At that time, there had been various special theoretical and experimental treatments of the distributions of mass in the sea, but supported by inadequate field observations. It was not until the relatively recent compilations of basic temperature and salinity observations and their subsequent analyses, that definite progress has been made in the interpretation of the ocean environment. Direct observations of the phenomena are essential to indicate the direction of theoretical research and to bring out new fields of investigation which ultimately become of practical importance.

Our present knowledge of sea surface wave characteristics is far from adequate to solve problems of practical significance. For instance, reliable witnesses have on occasions reported seas as being in excess of fifty feet in height from crest to trough, and the literature suggests fantastic sea conditions in the regions of tropical storms. On the other side of the ledger, it is of interest to note that the highest wave measured by the German research ship "METEOR," during three continuous years in the South Atlantic ocean, was about thirty feet.¹ These and other scientific observations suggest that reported wave heights are frequently exaggerated. The visual observations of individuals under duress do not have a high degree of reliability and the combination of wave frequency, steepness, and height contribute to an illusion of height alone. For instance, the coastal seas of New

* Contribution No. 435 Woods Hole Oceanographic Institution.

England are notorious for roughness conditions during the fall and winter seasons. The wave heights automatically recorded during the autumn of 1946 show that wave heights are not as great as may be inferred, and that combinations of wave height and wave frequency produce roughness conditions rather than wave height alone. Analyses of sea surface conditions in the vicinity of Woods Hole, during the autumn of 1946, show the most frequently occurring wave height to be of less than one foot, and an average wave height of 2.3 feet. Fifty per cent of the time the highest waves occurring in twenty minute sequences were less than three feet, and ninety-five per cent of the time less than fifteen feet.

Sea Surface Roughness and Military Operations

The absence of concrete data on the physical characteristics of the sea surface have resulted in a general inability to translate sea-surface conditions into operational or design situations, and a vagueness exists as to the limits of sea-surface conditions for operating various types of surface craft, as well as for other phases of maritime activity. In particular, this was brought into prominence by amphibious operations of the recent war.² Perhaps a clear idea of the situation may be brought out by brief reference to the invasion of the Normandy beaches in June 1944. The heights of waves on the assault beaches had been forecasted on board the Command ship (Ancon), for two days previous to the assault on June 6, and were transmitted directly to the staff of First U. S. Army. Observers on the beach at D-day verified that maximum wave heights in the off-shore debarking areas, and of surf on the target beaches, did not, in general, exceed 4 feet, as forecasted. Hence, the difficulties in carrying out the invasion plan because of adverse sea conditions were not the result of inadequate or erroneous advance information, but rather an inability to associate sea conditions to the details of the plan itself. In particular, sea conditions interfered with the timing of transfer of men and materiel from transports to small assault craft in the off-shore debarking areas, and in the eventual time and locations of landings on the target beaches. The result was that the neatly prepared paper operational schedule broke down in practice. It is known that, for instance, infantry units committed to build up firing lines on the beaches actually arrived off schedule and took shelter among the enemy beach obstacles, in process of being cleared by engineer demolition teams. The cost in loss of life and materiel was excessive.

A second example of a catastrophe during this military operation, which resulted from lack of adequate design to meet sea conditions, concerns the destruction of the artificial harbors and causeways off the Normandy coast during the storm June 19-20, 1944. These artificial structures had been erected, at great difficulty and expense, for the purpose of providing shelter for small surface craft engaged in the transferring of cargo from ship to shore during the critical period of the operation. However, it so happened that

two weeks after the initial assault, and at a time when the supply of American forces was critical, the installation was thoroughly destroyed by a storm. The loss of these structures seriously handicapped the continental military operations.

It has become apparent to even the uninitiated that inadequate information on surface conditions played a costly role in amphibious operations of the recent war. Reports of military operations in coastal areas, now almost invariably decry the lack of adequate prelanding information, and, in general, the effects of sea-surface conditions on all phases of the operation had to be learned in the field—usually under adverse conditions.³

From June to November of 1944, at the time cargo and personnel were being transferred from ships anchored off shore to the open Normandy beaches, a local state of the sea forecasting service was in operation under the direction of the European Theater Military Oceanographer.⁴ Some time after cargo operations had become stabilized, it was evident that the forecasting of impending sea conditions could be prepared in terms of the actual operation, that is, in terms of the feasibility of DUKWS and coasters to operate from transports to shore.

FIGURE 1 illustrates the actual relationship of maximum cargo unloadings in relation to height of surf on the Normandy beaches during October, 1944. By using maximum values, tonnage drops resulting from delays in assembling craft and other uncontrollable factors are eliminated. During October, the available arrangement and equipment was such that, under favorable conditions, a maximum of 10,000 long tons could be discharged daily. Thus, FIGURE 1 shows that surf heights up to one and one-half feet did not interfere with this maximum capacity. With greater surf height, cargo operations became difficult and for surf heights of three and a half feet the unloadings were reduced to about fifty-five per cent of the possible total. Surf heights between three and a half and four feet were critical. For the latter, total cargo unloadings of only twenty per cent efficiency were possible. Above four feet, operations were impaired so severely that the high maintenance costs were justified only in cases where the supply of inland armies was critical. At six feet surf height, all cargo operations were forced to cease regardless of demand.

The above simple relationship represents one of those things which, if it had been available during the planning stages of the Normandy operation, would have eliminated much of the guesswork and confusion regarding the supply of our inland armies. As it was, the initial elaborate plans for restoring ports and harbors were later dismissed in favor of maintaining cargo operations across the original Normandy beaches until the beginning of winter—something not thought possible before the invasion.

These brief references to the military aspects of this picture are given

* This author, Lt. Col., Corps of Engineer Reserve. Then commanding 6812 Engineer Technical Intelligence Team.

because of the writer's practical acquaintance with the examples and because the results drive home the need for investigation of sea-surface characteristics. When human lives are at stake, a high priority should be given to the investigation of those phenomena which have so clearly been demonstrated as controlling factors.

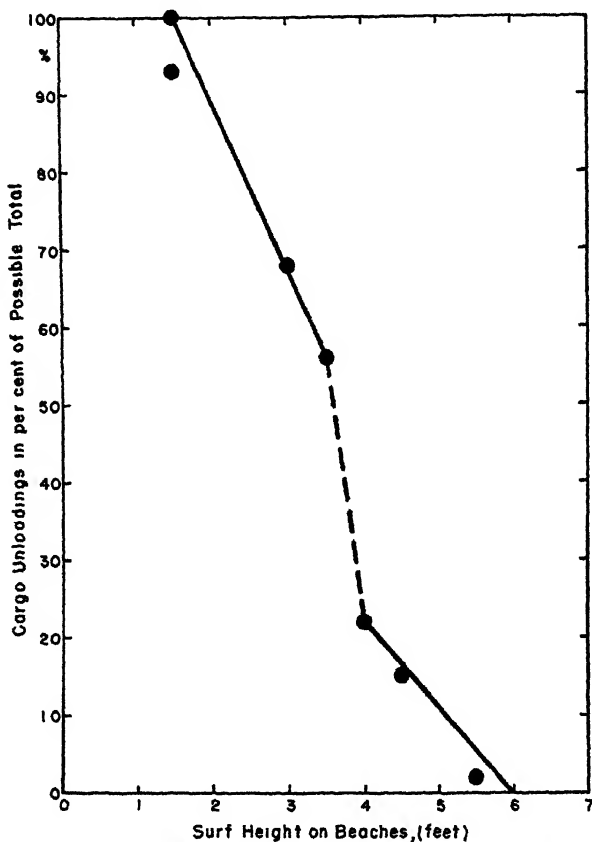


FIGURE 1 Observed relationships between surf heights and maximum daily tonnage discharged, Omaha Beach, Normandy, France, October 1 to 28, 1944. Drawn from data in Figure 2 "Military Oceanography in World War II".

Methods for Measuring Sea Surface Roughness Conditions

Roughness conditions of the sea surface may be measured by three general methods, all of which were developed as a result of wartime oceanographic research. These are: (1) By matched pairs of overlapping vertical photographs projected in stereo; (2) By evaluation of the state of the sea surface from wind patterns of past weather maps; (3) By automatic recording of pressure variations near the sea bottom.

The first method requires the use of specially equipped aircraft, including a mounted continuous strip Sonne Stereo Camera and a radio altimeter, to be operated by a trained pilot. Skilled interpretations of the resulting photographs are also necessary. This method, used with varying degrees of success during the war, served as an excellent check on trial forecasts of sea and swell.

The method of determining sea surface roughness conditions from wind distributions over the ocean is the simplest and most rapid approach to the problem⁴. It is limited, however, by the uncertainty of empirical relationships between wind and sea, as well as by uncertainties in computed wind distribution patterns over the open oceans. Later, when empirical relationships of wind and sea are better established, this method will probably be of top priority.

The method of evaluating sea-surface conditions from automatic recordings of pressure variations near the sea bottom has been in use at the Woods Hole Oceanographic and other institutions. The instruments in use are discussed by the builder A. A. Klebba.⁵ In my opinion, this method is, for the present, the most satisfactory available to provide urgently required basic data. Theoretically, the method is limited to the recording of pressure pulses of surface waves whose lengths exceed approximately twice the depth of instrumentation. However, regardless of this, the judicious installation of automatic underwater pressure recorders should have a high priority in that they are the only present means for providing state-of-the-sea-surface data under all conditions.

To the above methods should be added one involving photographic recording of wave heights against a graduated pole, either anchored to the sea bottom or in the deep ocean, held in place by means of a deep baffle extending toward the downward limit of wave action. Motion-picture recordings of the wave profiles are later scaled for wave height and wave period. This method has been successfully used in water up to 125 feet deep by this institution, and during the present cruise of the ATLANTIS wave recordings are being made by the above deep-water method. The exact details of the ATLANTIS apparatus are not known, as it has undergone some modifications while at sea. Essentially, it consists of a 30 foot pole, graduated in alternate black and white stripes, to which a long baffle is attached. After the pole is set adrift, timed photographic recordings on 16-millimeter film are made for 90-second intervals. Some recent results, obtained from the recordings in the North Atlantic are illustrated by FIGURE 2. These data will be later analyzed with reference to prevailing wind distributions.

Some Significant Results of Sea Surface Roughness Measurements

The remainder of this paper is concerned with certain analytical results of sea surface wave data obtained from evaluations of underwater pressure instrument records, in order to review briefly some pertinent results to be

obtained from this type of information. The pressure-measuring elements were located, one off Cuttyhunk Island at a depth of 75 feet, and the other off Bermuda, at a depth of 120 feet (FIGURE 3). In both cases, they were connected to onshore recorders several miles away. Because of installation locations, pressure pulses of wave lengths less than 150 feet at Cuttyhunk and less than 240 feet at Bermuda were not expected to be recorded. The operational schedule of the Cuttyhunk instrument, from July, 1946, to May,

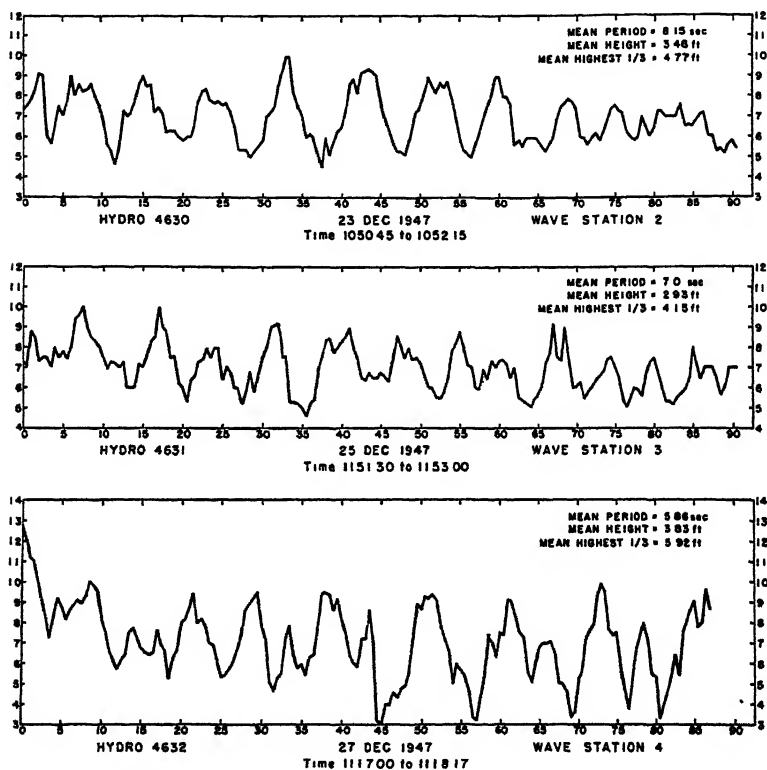


FIGURE 2. Results of measurements of sea surface roughness in the North Atlantic during ATLANTIS winter cruise, December, 1947.

1947, was a continuous 20-minute recording every 6 hours; that at Bermuda, from February, 1947, to May, 1947, a continuous 20-minute recording every 2 hours. The records (FIGURE 4) were scaled for pressure wave heights and evaluated in terms of wave heights at the physical sea surface immediately overhead.

Because with surface waves the downward acceleration of the water particles below the crest counteracts the effect of gravity, and the upward acceleration of those below the trough adds to it, pressures are not perceptible at depths where orbital motion is imperceptible, regardless of surface wave

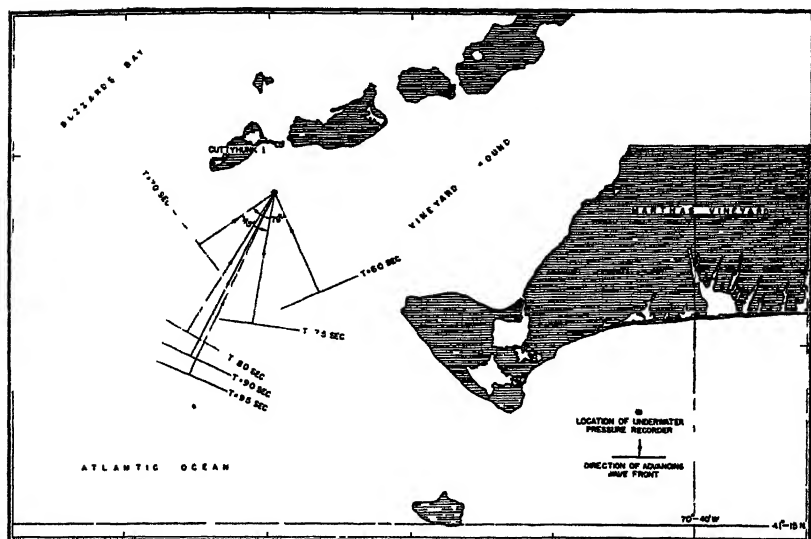


FIGURE 3a

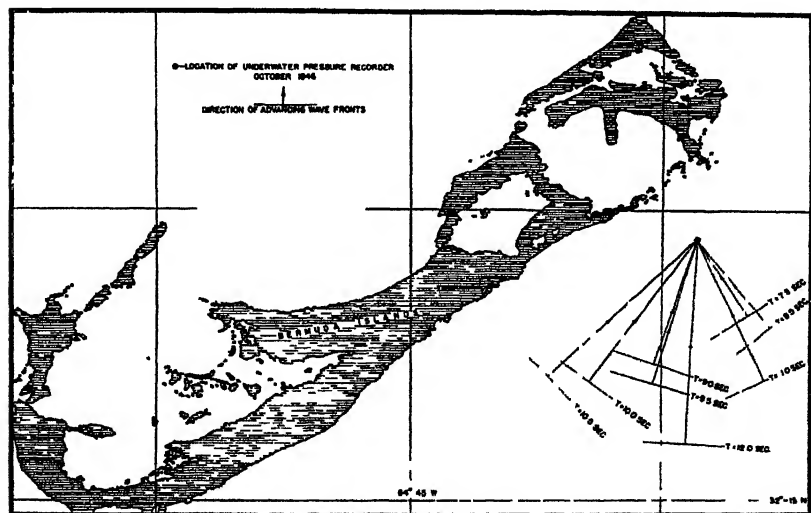


FIGURE 3b

Locations of underwater pressure wave recorders at Cuttyhunk (3a) and Bermuda (3b). Directions of advancing surface wave fronts computed from differences in surface and bottom phase angles of wave components, as obtained from experimental observations on 27 June, 1946 at Cuttyhunk, and 25 October, 1946 at Bermuda.⁵

heights. Theoretically, the relation between surface (ΔP_s) and bottom (ΔP_h) pressure fluctuations (at depth h) is given by:⁶

$$\frac{\Delta P_h}{\Delta P_s} =$$

where

$$k = \frac{2\pi}{\text{wave length}}.$$

The damping phenomenon is, thus, selective to the extent that longer period waves, frequently obscured at the sea surface, are generally recognizable in underwater pressure records. By the same token, shorter wave lengths, which do not exceed approximately twice the depth of instrumentation, are not recorded, regardless of their height. This fact calls for a judicious selection of instrument sites, when measuring surface waves by the method of bottom pressure fluctuations.

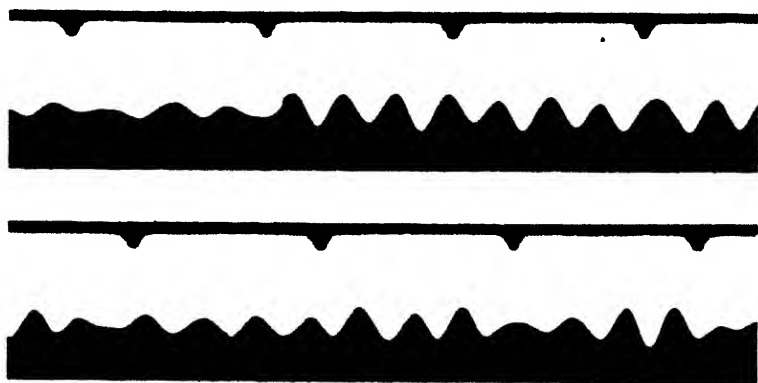


FIGURE 4. Reproductions from parts of two underwater pressure wave records obtained from the Cuttyhunk location on 14 June, 1946. Time ticks are for thirty-second intervals. Bottom wave heights computed from instrument factors and evaluated in terms of surface wave heights as described in text.

The results given by the above equation are not in complete agreement with observations, hence the possibility must not be overlooked that other solutions may be discovered to satisfy the observed conditions. With this in view, a series of experiments was performed at both Bermuda and Cuttyhunk, whereby wave heights of the physical sea surface immediately overhead the underwater pressure records were photographically recorded for comparison with wave heights computed from instrument recordings by theory. It was found that, for both localities, better representations of sea-surface conditions were obtained when theoretical results were multiplied by a factor of 1.35. This factor is accepted only for the localities involved. It is doubtful, in view of the complex variations of natural conditions, if a universal factor is possible and, for the present, empirical relationships need be established for each locality from experiments in the field. Eventually, with accumulation of field results, other solutions will be discovered to modify existing wave theory so as to better satisfy observed conditions.

In this connection, it is of interest to point out that many instances in the

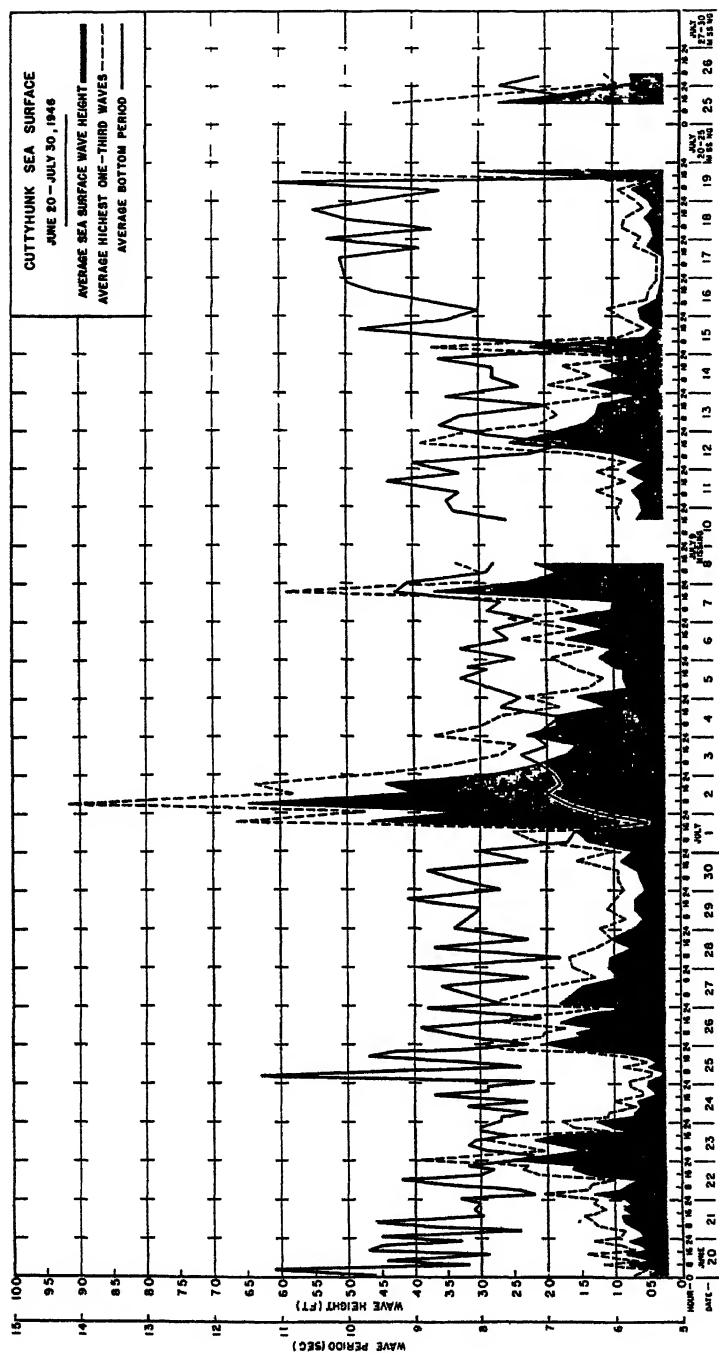


FIGURE 5. The state of the sea surface off Cuttyhunk Island, June 20 to July 30, 1946, as computed from six-hourly underwater pressure recordings.

literature show observed conditions of wave action in coastal areas do not fit within the framework of existing shallow-water wave theory. During the war, it was particularly apparent to those engaged in producing reconnaissance intelligence that theoretical reductions of wave observations were not compatible with existing conditions. A specific example is the determination of underwater depths from measured changes in the wave patterns of aerial photographs. Almost invariably, it was necessary to introduce empirical factors to take up discrepancies between computed and theoretical

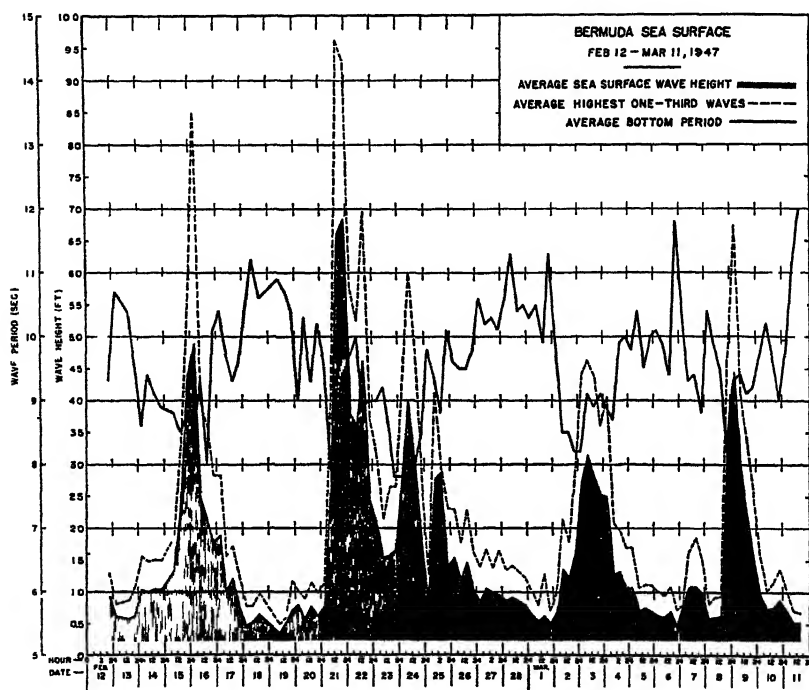


FIGURE 6. The state of the sea surface off Bermuda, February 12 to March 11, 1947, as computed from two-hourly underwater pressure recordings.

values. However, the basis for the revision of present wave theory and for the discovery of more suitable solutions depends on obtaining accurate observations on the state of the sea surface itself.

Examples of sea surface wave characteristics⁶ at Cuttyhunk from June 20 to July 30, 1946, and for Bermuda from February 12 to March 11, 1947, are illustrated by FIGURES 5 and 6. In these figures, average six-hourly sea surface wave heights, average highest one-third waves, and average bottom periods are shown. These data, when arranged in suitable frequency distributions and subjected to standard statistical procedures, provide information on sea-surface characteristics at the two localities considered. The

prognostic usefulness is dependent on the extent to which the data are characteristic of usual prevailing conditions.

TABLE 1 summarizes the seasonal statistical properties of sea surface roughness conditions at Cuttyhunk and Bermuda, based on analyses of more than 1600 twenty-minute records. The data show that seasonal changes in the state of the sea are reflected in the extreme, rather than in the mean conditions. Thus, at Cuttyhunk, while the mean wave height increased from 1.3 feet in summer to 2.3 feet in autumn, the heights of the highest waves increased from 16 to 22 feet. At Bermuda, the situation was similar. Mean wave height diminished from 1.7 feet in winter to 0.8 feet in spring,

TABLE 1
STATISTICAL SUMMARY OF SEASONAL SEA SURFACE PROPERTIES FOR CUTTYHUNK
AND BERMUDA*

Season	Mean Wave Height					Maximum Wave Height			Mean Wave Period		
	Mode	P.E.	Occurrence below mode	Variability		50 per cent occurrence	95 per cent occurrence	Highest wave	Mean	Mode	P.E.
<i>Cuttyhunk Wave Station</i>											
Spring	1.3	0.6	0.1	41	.81	to 2.8	to 9.0	16	8.6	7.9	0.2
Summer	1.3	0.7	0.1	48	.80	to 2.5	to 9.5	15	8.4	(8.3)	0.1
Autumn	2.3	0.9	0.1	39	.83	to 2.9	to 14.5	22	8.1	6.8	0.1
<i>Bermuda Wave Station</i>											
Winter	1.7	0.8	0.1	42	0.81	to 2.8	to 10.5	17	9.6	9.5	0.1
Spring	0.8	0.7	0.02	62	0.52	to 2.3	to 4.5	7	10.1	9.3	0.1

* Statistical summary of seasonal characteristics of sea surface roughness at Cuttyhunk and Bermuda, computed from recording* of underwater pressure fluctuations. Wave heights in feet, periods in seconds; occurrence figures in per cent.

and the maximum wave heights from 17 to 7 feet. Results such as these permit estimates of expected regional operational conditions.

The Growth and Decay of the Sea Surface

An ideal solution to the practical problem of sea surface roughness would be one establishing statistical relationships between rates of change of sea surface wave heights and wave periods and the prevailing state of the sea. The data available do not permit this extensive analysis, in that the relatively few months comprise only single series of observations. Consequently, our results are valid only in so far as the single series of observations conform to near normal values for the months in question.

The rates of change of wave heights were computed from the Cuttyhunk six-hourly and the Bermuda two-hourly records. In treating the general case, frequency distributions for the separate positive and negative rates of

change at each locality were formed. The statistical quantities describing them are tabulated in TABLE 2.

The situation demonstrated by TABLE 2 indicates a remarkable uniformity in rates of growth and decay of sea-surface roughness. Both are about equal, with an overall range up to 1.9 feet per hour. The most frequently occurring changes in sea surface wave heights are very small, about 0.05 feet per hour, occurring about 30 per cent of the time. Rates of change of sea surface wave heights occurring 50 per cent of the time do not exceed 0.12 feet per hour, and 75 per cent of the time do not exceed 0.25 feet per hour. At both localities, sea surface wave heights increased and diminished at approximately the same average rate, namely 0.17 feet per hour. The rates of change of sea surface wave height in excess of one foot per hour occur less than 2 per cent of the time. It is apparent that only in exceptional cases does the state of the sea surface roughness change rapidly.

TABLE 2*

	Cuttyhunk		Bermuda	
	$+\frac{\Delta H}{\Delta t}$	$-\frac{\Delta H}{\Delta t}$	$+\frac{\Delta H}{\Delta t}$	$-\frac{\Delta H}{\Delta t}$
Mean	0.16	0.14	0.18	0.20
Mode	0.05	0.05	0.05	0.05
Probable error	0.01	0.01	0.01	0.01
50 per cent occurrence	to 0.08	to 0.08	to 0.10	to 0.12
75 per cent occurrence	to 0.19	to 0.17	to 0.20	to 0.25
Maximum value	1.5	1.5	1.9	1.9

* Statistical summary of rates of change of wave height ($+\Delta H/\Delta t$ = increasing; $-\Delta H/\Delta t$ = decreasing, feet per hour) for Cuttyhunk (July 1946 to May 1947) and Bermuda (February 1947 to May 1947).

Interrelationships of Wave Characteristics

A study of the data has brought out that, at both Cuttyhunk and Bermuda, higher wave heights are generally accompanied by lower wave periods and vice versa. Thus, at the former locality TABLE 1, we see that increased mean wave heights from 1.3 feet in spring to 2.3 feet in autumn is accompanied by diminished mean wave periods, from 8.6 to 8.1 seconds. Similarly, at Bermuda, where mean wave height diminished from 1.7 feet in winter to 0.8 feet in spring, the accompanying mean wave period increased from 9.6 to 10.1 seconds. This is the observed overall pattern. With the exception of occasional departures, higher surface wave heights and wave periods are similarly related. Thus, FIGURE 7 shows that relationships between rates of change in wave height and wave period, as derived by least square fits, are as follows:

$$\text{for Cuttyhunk} \quad \frac{\Delta H}{\Delta t} = -0.42 \frac{\Delta T}{\Delta t},$$

$$\text{for Bermuda} \quad \frac{\Delta H}{\Delta t} = -0.26 \frac{\Delta T}{\Delta t}.$$

The Determination of Operational Wave Heights from Average Wave Heights

The situations discussed so far are based on mean wave heights for 2-minute intervals every 2 or 6 hours, as the case may be. Coincident with the computation of mean wave heights, the means of the highest one-third surface waves were also determined for each record. These latter values are termed "operational wave heights," in that they better represent practical effects of sea surface roughness conditions than do mean values for all wave heights. The deduction is arbitrary, but from the standpoint of practical operations a somewhat better index of sea surface roughness conditions is provided.

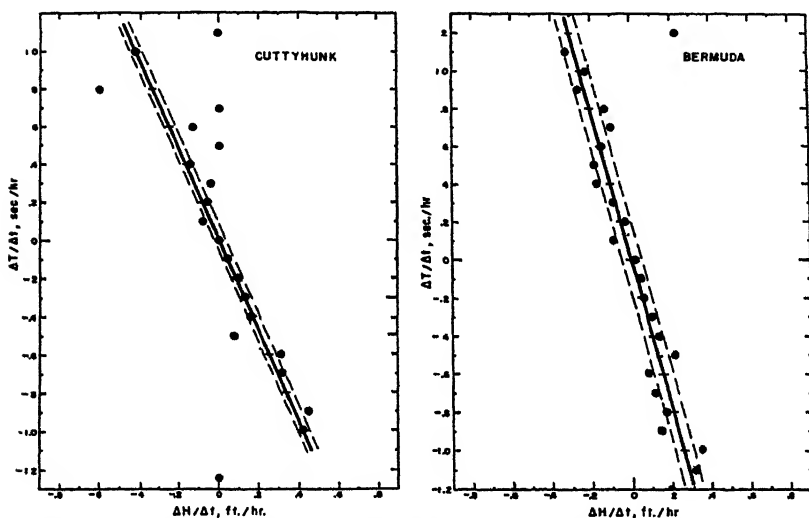


FIGURE 7. Relationships between the rates of change of wave periods and wave heights for Cuttyhunk and Bermuda.

The data from both Bermuda and Cuttyhunk show a significant relationship between the average of all wave heights and the averages of the highest one-third wave heights. This is illustrated for Cuttyhunk (639 observations) and for Bermuda (1022 observations) by FIGURE 8. Independent fits of straight lines by least squares to both sets of data give identical results, thus:

$$\text{Mean of highest one-third waves} = 1.57 \times \text{mean wave height.}$$

Hence, it is reasonable to compute operational wave heights by multiplying the mean of all wave heights by the factor 1.57. However, as for other factors derived in this study, they are, for the present, considered pertinent only for the localities for which derived. It is considered too early to generalize on the meager results of this and similar investigations.

These and other interrelationships of sea surface wave characteristics

contribute to an understanding of the complex sea surface patterns, as well as indicating regularities of practical significance within the patterns themselves. The few examples demonstrate that adequate field observations of the sea surface, when subjected to statistical analyses, will provide answers to many practical questions, as well as point the way to one phase of theoretical research on the sea surface.

As a final example, a functional method for evaluating operational sea surface conditions for any locality and for contrasting these conditions over

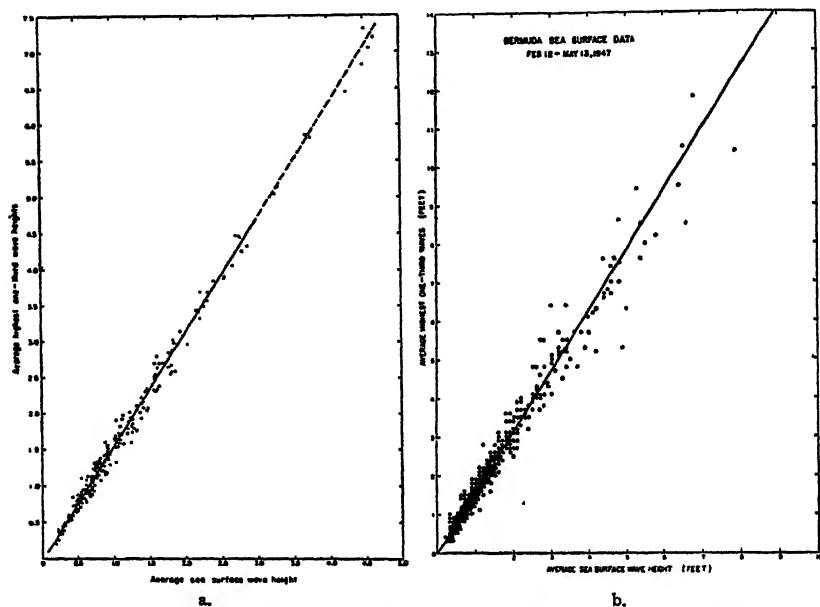


FIGURE 8. Relationship of average sea surface wave heights and average heights of highest one-third surface waves for Cuttyhunk (8a) and Bermuda (8b). The independent fits of straight lines by least squares to both sets of data give identical results, thus: *Means of highest one-third waves* = $1.37 \times$ *mean wave height*.

the world's oceans is outlined. Since its development depends on available sea surface data its present use is restricted to a few oceanic localities.

From a practical standpoint, operational conditions of the sea surface may be considered as a function of wave height and wave period. Thus, in regions of permanently restricted fetch (as the English Channel), there is far less variability in operational conditions than in an area such as the Eastern American coast, where all wave frequencies within the wind generating band are possible. A representation of operational conditions by combinations of wave height and wave period for the waters in the vicinity of Woods Hole, during the summer of 1946, is shown by FIGURE 9. In this example, the various combinations of operational wave heights and wave periods into favorable, unfavorable, and conditional situations are with reference to small-boat work. The division is arbitrary and presented pri-

marily from the standpoint of a method. The boundaries of wave height-wave period combinations may be selected with reference to any type of operation on the basis of experience. Each figure in Part A, FIGURE 9, represents the per cent occurrence of the wave height-wave period combina-

PART A

WAVE HEIGHT	WAVE PERIODS							PERCENT SUMMARY
	5-6	6-7	7-8	8-9	9-10	10-11	11-12	
10-11	0.4							0.4
9-10								
8-9	0.4							0.4
7-8	0.9	1.8	0.4					3.1
6-7		0.8						0.8
5-6	0.5	1.8				0.4		2.7
4-5	0.4	2.3	1.7					4.4
3-4	0.4	1.4	4.9	0.9	0.5			8.1
2-3		3.5	10.7	3.6	0.4			18.2
1-2		1.8	18.7	12.5	3.6	0.4		37.0
0-1			4.4	8.4	8.9	1.8	1.5	24.8
PERCENT SUMMARY	2.6	13.8	40.8	25.4	13.3	2.2	1.5	100

Summary Sea Surface operating conditions in Percent of Time, Cuttyhunk, Summer 1946. Wave height (feet) = average highest one third wave during twenty minute intervals. Wave period (seconds) = average recorded at bottom. Small boat operating conditions: Favorable; bounded by solid lines; unfavorable; bounded by dashed lines.

PART B

LOCATION	SEASON	SEA SURFACE CONDITION		
		Favorable	Unfavorable	Intermediate
CUTTYHUNK	Spring	73	13	14
	Summer	77	12	11
	Autumn	55	31	14
BERMUDA	Winter	74	11	15
	Spring	90	1	9

Summary Sea Surface seasonal operating conditions in Percent of Time at Cuttyhunk and Bermuda as computed from boundaries established in Part A.

FIGURE 9.

tion. Summing values within these arbitrarily selected zones gives the following estimate of summer operating conditions: favorable, unfavorable, and intermediate operating conditions, 77, 12, and 11 per cent, respectively.

The extension of this scheme to the summer and autumn seasons for Cuttyhunk and the winter and spring seasons for Bermuda, produces the seasonal operational summary of Part B, FIGURE 9. This enables a comparison

of seasonal operational conditions, within defined boundaries. It shows that, for the Woods Hole region, the 75 per cent favorable operating conditions of spring and summer are reduced to about 50 per cent during the autumn. At Bermuda, the 75 per cent favorable operating conditions in winter increase to 90 per cent in spring. If the data were available, similar schedules for any type of marine operation could be established throughout the world, and much of the guesswork now characterizing operations of this type would be eliminated.

TABLE 3*

	<i>Cuttyhunk</i>			<i>Bermuda</i>	
	<i>Spring</i>	<i>Summer</i>	<i>Autumn</i>	<i>Winter</i>	<i>Spring</i>
Wave height					
Mean	2.0	2.2	3.7	2.6	1.3
Mode	0.9	1.2	1.4	1.2	1.1
Maximum	16	15	22	17	7
Wave length					
Mean	381	365	337	472	520
Mode	325	360	233	461	444
Maximum	832	900	832	1040	1040
Wave velocity					
Mean	26.0	25.5	24.5	29.1	30.6
Mode	24.0	25.3	20.5	28.8	28.2
Maximum	38.6	40.2	38.6	43.2	43.2
Wave steepness					
Mean	.54	.65	1.16	.55	.25
Mode	.28	.34	.61	.27	.25
Maximum	8.7	8.2	12.0	9.2	3.8

SUMMARY OF SEASONAL OPERATING PROPERTIES FOR CUTTYHUNK AND BERMUDA COMPUTED FROM VESSEL OBSERVATIONAL DATA. Wave heights, (feet) = averages of highest one-third waves. Wave length (feet) = $5.12 \times (\text{period})^2$. Wave velocity (knots) = $3.03 \times (\text{period})$. Wave steepness (percent) = wave height/wave length; maximum values computed for six-second periods.

Sea Surface Roughness Factors in Relation to Design

So far, this discussion has been concerned with operational results of sea-surface analyses. However, the same basic data from which they have been derived also lend themselves to computations of the forces of waves and inertia exerted on the hulls of surface craft and fixed marine structures. These computations are of practical importance to the designer of surface craft and marine structures, although a closer coordination between the oceanographer and the marine designer is needed to bring out their full significance. There is a fundamental difference in the analyses of data, in that, from an operation standpoint, emphases are placed on the mean or average conditions, whereas for design, the maximum or extreme conditions play the prominent role.

A computation summary of some relevant sea-surface properties significant to marine structure design is given in TABLE 3. They serve as examples

of what may be computed from the basic data. For the region under consideration, it is of interest that the shorter and higher seas of fall and winter produce greater steepness values than at other times of the year. The maximum steepness values, although occurring rarely, are, for waves of 800 to 1000 feet long, greater than generally allowed for in the stress computations of marine craft design. The maximum wave lengths recorded are about three times greater than those most usually occurring.

Conclusions

Although observations on oceanic phenomena have been recorded since the dawn of history, the systematic investigation of the oceans is a young science. Some of the branches of oceanography, such as coastal tides and tidal currents have, because of their practical importance, been placed on a systematic basis years ago, and, as a consequence, are further advanced in both theory and practice. As far as sea surface waves are concerned, the numerous visual and often sporadic observations made from ships at sea have been useful for past objectives in providing pictures of general situations, but are for the most part not usable for the precise analyses now required for present operational and design trends. This was realized during the war, when the practical problems of military oceanography called for development of special wave recording instruments, as well as for special wave studies. With this impetus given to wave research, we have come a long way in our knowledge of sea-surface characteristics, and are in a position to formulate new lines of research pertinent to both scientific investigation and industry. In addition, there is the benefit to problems arising in connection with the future defense of this country, it being significant that the historical inference of security from oceanic isolation is now no longer correct. The environmental problems of new types of coastal installations and the maintenance of world-wide lines of rapid communication add to the necessity for a program aimed at understanding the characteristics of the sea-surface layers. The knowledge gained from investigations of the last war is no longer adequate to present peacetime problems of either science or industry, because neither has stagnated.

The practical application of results of wave research and the expansion of this phase of oceanography call for a closer coordination and cooperation between the oceanographer and the engineers and production personnel of industry. The separation in duties of the personnel involved requires that industry be advised by oceanographers with backgrounds of research, so they can envision future lines of research and not permit a stagnation at present levels. To keep abreast of an advancing science, the expansion of present oceanographic research facilities and personnel, such as can be accomplished by cooperation with industry, is essential. This requires bringing to the attention of industry such of the numerous present and future uses of results of oceanographic research as benefit their economy.

Sea surface wave research is an oceanographic subject which has been brought into the limelight as of great practical importance. It now remains to formulate adequately the scientific and practical problems to support a continuing progress. This paper has been concerned only with the practical analyses of results of field observations. From this viewpoint, the primary requirement is establishment, in the not too distant future, of a controlled program of field observations, located at strategic places throughout the world's oceans. As the results of such analyses become available, we approach a position enabling us to forecast and provide prognosis for areas where direct observations are obtained only with extreme difficulty, and where, as has been demonstrated in the past, they will undoubtedly be required in the event of another world emergency.

Bibliography

1. SCHUMACHER, A. 1927. Ozeanographische Sonderuntersuchungen Zeitschrift der Gesellschaft für Erdkunde zu Berlin. 5/6: 304-307.
2. SEIWELL, H. R. 1947. "Military Oceanography in World War II". The Military Engineer. 39(259): 202-210.
3. 1945. Omaha Beachhead Historical Division, U. S. War Department.
4. SVERDRUP H. U. & W. H. MUNK, 1943. Wind waves and swell Principles in forecasting. U. S. Navy Hydro. Off. Misc. Pub. 11: 275.
5. KLEBBA, A. A. 1949. Details of Shore Based Wave Recorder and Ocean Wave Analyzer. Ann. N. Y. Acad. Sci. 51(3): 533-534.
6. SEIWELL, H. R. Results of Research on Surface Waves of The Western North Atlantic. (I). Papers in Physical Oceanography and Meteorology. 10(4).

Discussion of the Paper

DR. KENNETH S. M. DAVIDSON (*Director, Experimental Towing Tank, Stevens Institute of Technology, Hoboken, N. J.*):

I am interested in the seakeeping qualities of ships. This is a broad subject, and one on which there exists little authoritative information, despite its great age.

It has become increasingly clear to me that much depends upon establishing close ties between research in oceanography and in naval architecture. The naval architect needs to know more about the nature of the sea surface to which ships have to accommodate themselves, and the oceanographer needs to know more about the sea surface for many other reasons as well. It is natural, therefore, that the two should combine forces. They are feeling their way at the present time, and are in the very early exploratory stages of cooperation. The naval architect does not quite know what questions to ask the oceanographer, and the oceanographer does not quite know what questions to ask the naval architect. They do not yet know how to cooperate to mutual advantage.

The naval architect still represents rough water by a single train of regular trochoidal waves. This simplification has worked out fairly well, and, to the extent that there is usually some sort of "predominant" wave train in any actual seaway, as I gather the oceanographers admit, seems to be justified as a first approximation. In any event, it has been put to good use in

studying such things as, for example, the rolling of ships in beam seas. On the other hand, it is obviously inadequate as a basis for a full understanding of even a simple motion of this sort. Consider: the damping in roll of ordinary ships not equipped with mechanical stabilizing devices corresponds to amplification ratios (roll amplitude/maximum wave slope), at resonance, that lie within a range of something like 4 to 20. Now, there is no evidence that an amplification ratio of the order of 20 means an unsafe ship at sea, as would surely be the case if resonant wave encounters persisted for sufficient time to allow the amplitude in roll to build up. Neither is there good quantitative evidence of improvement (or otherwise), if the amplification ratio is materially reduced, by increasing the damping. Evidently, then, the irregularities of actual sea waves must play a big part in determining the rolling behavior of ships.

This is not a very profound conclusion, perhaps, and it has certainly been suspected many times. The question is what might conceivably be done about it. One answer that suggests itself readily enough is to undertake a much more searching dynamic analysis of rolling than has yet been attempted, employing detailed contour maps of the sea surface, perhaps like those of Schumacher (Meteor Expedition). One would want to be sure of using "typical" maps and to have a clear view at the start of their relation to representative weather conditions. This means that the oceanographer probably ought to select them as well as provide them.

There are, of course, other ways of studying rolling, and I do not wish to be understood as recommending this one especially. The point I want to make, is simply, that if anything new is to be learned about rolling, the irregularities of actual sea waves will almost certainly need to be taken into account. There is no doubt in my mind that this is equally true of all other motions of ships in a seaway, as well as of the forces and stresses imposed on the ship structure.

I have dwelt upon the need of the naval architect for detailed information regarding the character of the sea surface in the immediate vicinity of a ship, because recent oceanographic studies seem to have emphasized mainly the overall aspects, like mean dimensions of waves, probabilities of occurrence, etc., and to have paid little attention to detailed aspects like wave configurations. I do not mean to imply that the naval architect is not interested in the overall aspects. He is, but he needs something more. FIGURE 1 of Mr. Seiwell's paper poses the question fairly. By presumption, the curve on this chart will be shifted bodily toward the right as designs are improved. To improve designs will not be easy at best, but it will be much harder unless a better knowledge is gained of operating conditions, in considerable detail.

There is much to be done. As matters stand at present, a free interchange of views between research groups in naval architecture and in oceanography is probably the most important single step that can be taken.

A COMPARISON BETWEEN RECORDED AND FORECAST WAVES ON THE PACIFIC COAST*

By JOHN D. ISAACS AND THORNDIKE SAVILLE, JR.

University of California, Berkeley, California

History is replete with records of damage and havoc wrought by waves generated by storms at sea. The affairs of men and nations have often been affected by the loss of ships or entire fleets. In ancient times, Xerxes's fleet of 300 ships and 20,000 of his men were lost in a storm off Mount Athos. Shortly afterward, his bridge of ships across the Hellespont was destroyed by storm waves, and Xerxes ordered that the sea be given 300 lashes. The entire course of history was affected by the scattering and partial wrecking of the great Spanish Armada in a storm. More recently, the efforts of the Allied forces were seriously hampered when a part of the preparations for the Normandy beachhead was destroyed by sea waves. The lives and thoughts of many men have been devoted to the study of the vagaries of the sea, to the calm and violent moods of sea and surf. Great concentration of investigation was given to wave theory and the prediction of wave conditions during the war, when a knowledge of sea conditions was essentially indispensable for the success of amphibious landings and the transportation of forces to frontal zones. In peace, however, the importance is as great. In times of heavy seas, ocean commerce is seriously disrupted, many harbor entrances become impassable, fishing operations are suspended and shore installations, piers, homes, and breakwaters are frequently destroyed. These heavy seas do not necessarily accompany local weather disturbances. Thus, their occurrence is not indicated in local weather forecasts, and they may arrive without warning. Navigators have, therefore, had no opportunity to shape their courses and arrive at havens before the entrances become impassable. R. S. Arthur has pointed out that the destructive storm waves which reached the islands of the Hawaiian group on January 2-5, 1947, and which caused damage estimated at between one and two million dollars, could have been forecast 24 hours in advance, and much destruction prevented. Aside from the immediate function of warning, these forecasting techniques become a powerful tool in statistically determining the probable state of the sea at any point where adequate historical weather maps exist over the adjacent bodies of water. This has great application in the design of engineering structures, dams in inland regions, breakwaters, piers, groins, and seawalls, in the selection of transoceanic air routes, and in the guidance of marine operations of all kinds. Hence, it is of great importance that the methods of predicting the state of the sea be perfected.

Various investigators have attacked the problem of relating wind and wind

* The work summarized in this paper was part of an extensive investigation on waves and surf for the Oceanographic Section, Bureau of Ships, United States Navy.

waves. Krummel in 1911, Cornish in 1912, Defant in 1929, Patton in 1932, and others have given surveys of the general knowledge of the characteristics of ocean waves. From these, various empirical formulae have been developed. Cornish gives the height of the highest waves as $H = 0.48U$ (where H is in meters, and the wind velocity, U , in cm/sec). Zimmerman gives $H = 0.44U$. Rosby and Montgomery have suggested the formula $H = \frac{0.3}{g} U^2$, which is dimensionally correct. Other similar empirical equations have been evolved relating wave velocity with wind velocity. Stevenson has related the maximum waves to fetch by the formula $H = \frac{1}{3} F$ (where H is in meters and fetch, F , in km). Stokes and Jeffreys have advanced theories for dealing with ocean waves.

It has been only recently, however, that the theory of forecasting of ocean waves has been put on a scientifically sound basis by Sverdrup and Munk. Sverdrup and Munk propound the presence of "significant" or "characteristic" waves: those having the average height and period of the one-third highest waves. Experience has indicated that these characteristic waves correspond most closely to those estimated by visual observations, and consequently, the present-day empirical data consist chiefly of these characteristic waves. It is known that energy is transmitted to the waves by the wind in two ways: first, by a tangential, frictional stress on the water surface; and second, by the normal pressure, giving a pressure differential between the windward (high pressure) and the leeward (low pressure) sides of the wave. It may be of incidental interest to note that a theory developed on the basis of energy exchange due solely to normal pressure indicates that the waves cannot reach velocities exceeding the wind velocity (one of the big drawbacks to most previous theories), whereas a theory which also takes into account the energy transfer by tangential stress does not place this restriction on the development. Predicating the existence of characteristic waves, it was possible to formulate a differential equation on the basis of the rate of energy transfer from the wind and the rate at which the wave energy advances, from which special solutions of the relationships between the wave characteristics and fetch, wind velocity, and duration were obtained. The numerical constants involved have been evaluated in terms of empirical data, and the results expressed in graphical form by Sverdrup and Munk. Very little, however, is known about the generation of waves by winds much above 60 knots, because of the paucity of data, and the graphs obtained apply only to winds below this velocity. In common with all other wave phenomena, when ocean waves advance into a region of calm (or decay), their period and velocity increase, while their amplitude decreases. The loss of energy due to air resistance may be computed, and, assuming the energy loss from tangential stress and viscosity to be negligible, curves showing the relationships between the wave characteristics, the decay distance, and initial wave characteristics may be drawn.

For a more complete description of the theory involved here, reference is

made to "Wind, Sea and Swell, Theory of Relations for Forecasting," by Sverdrup and Munk.*

Data obtained from limited fetch work by the University of California at Clear Lake, California, and more recently, from Abbotts Lagoon at Point Reyes, California, very strongly substantiate the theoretical relationships with minor corrections to empirical constants.

Forecasting of sea and swell for comparison with observations was initiated in 1944, by personnel of the University of California, but a confirmation of the forecasting relationships by reliable prototype measurements awaited adequate recorded data on the Northern Pacific Coast.

There are a number of methods of measuring wave heights and periods, and the reliability of the data obtained depends to a great extent upon the method employed. Visual shore observations can be made from piers, reading a staff gauge or using a plumb bob on a graduated line. An experienced observer can obtain good results by these methods. The wave, however, is influenced somewhat by the piling of the pier, and the change in the bottom induced by the presence of the pier. Breaker measurements can be made by optical aids or photogrammetric methods. The University of California field party has used both extensively. A doubt is introduced into the deduced deep-water condition, however, as the waves undergo some final refraction over temporary beach details. Observations offshore have been made by various instruments with varying success. These have taken the form of instruments which record the position of the water surface and those which record the pressure fluctuations at some depth. The method of measuring the pressure fluctuations below the surface possesses a unique advantage over that giving a trace of the water surface in that the effect of waves shorter than any particular period can be eliminated by appropriate depths of installation. Thus, it is possible to avoid the obscuring effect of small, local, wind waves and chop, and to record only the consequential waves which dominate the surf. Many types of pressure transmitters have been designed: variable resistance, capacity, selsyn linked, etc. The type employed in the shore wave recorders installed by the University of California utilized a potentiometric circuit. Two installations on the Pacific Coast have been in operation for almost a year and have proven very satisfactory.

The selection of a site for wave measurements requires the evaluation of a number of natural factors. The instrument should be placed at an exposed point on the coast so that all waves reach it, and reach it with as small an amount of refraction as possible. For this reason there should be no reefs or canyons to seaward of the instrument at moderate depths. A gently sloping beach is preferable to vertical cliffs to reduce the probability of reflected waves complicating the record. Installation depth of 60-70 feet should be reached within a reasonable distance from shore (3000 feet). The cable must pass through the surf zone in a relatively protected area, to permit

* Cf. references.

cable laying from small craft and so that longshore currents or surf will not move or damage the cable. Heceta Head, Oregon, and Point Sur, California, fitted these requisites well and were a fairly large distance (600 miles) apart (FIGURE 1). Shore stations already existed at these sites in which the recording unit could be placed. Future installations are now being planned at three more points along the coast—Quillayute, Washington, Point Cabrillo, and Point Arguello, California.

The actual installation was done by Navy DUKW. At Point Sur, a steel life raft was adapted to carry a spool of cable from the beach to sea, the cable unreeling as the raft was towed seaward. The cable was installed on the raft on shore, and the shoreward end secured to posts on the beach and later connected by lighter cable to the recorder. The pressure-transmitting instrument was attached to a metal triangle support, spliced to the cable at sea, and lowered to sea bottom from the DUKW at the desired location.

The wave recorder runs on slow speed (3 inches per hour) for 5 hours 35 minutes, and then on fast speed (3 inches per minute) for 25 minutes. The record can be analyzed for height at any time, but for period only when running on fast speed. Records are analyzed at 12-hour intervals, and at every intermediate point of possible interest, or where a definite change in wave characteristics occurs. The instruments were calibrated in a dynamic testing chamber before being installed, and graphical overlays constructed from this calibration and the known relationship of pressure fluctuation with depth, wave period, and height. The period of the waves is determined and the heights of the highest third of the major train of waves obtained from the overlay and averaged to give the characteristic height. The waves are almost invariably complex, consisting of several trains of waves of about the same period, probably generated in the same fetch, with, very often, the added complication of a secondary wave train from another fetch. In this case, both primary and secondary waves are measured. These analyses give the characteristic wave heights and periods at the depth of the instruments (65 feet at Point Sur, and 50 feet at Heceta Head), and these must be transferred to deep-water characteristics for comparison with the forecast values. Two effects enter into this reduction, that of refraction and that of the change in energy advance. For the former, refraction diagrams were drawn for both installations for waves of various periods and directions and the corresponding refraction coefficient determined. For the latter, the wave period was assumed constant, and the deep-water wave length, L_0 , taken as $5.12T^2$, from which d/L_0 was computed. H/H_0' then was taken from the appropriate graph, and the deep-water height computed as the quotient of the recorded height and these two terms.

$$\left[H_0 = \frac{H}{K \left(\frac{H}{H_0'} \right)} \right]$$

Some error occasionally is introduced into the results here, as the actual wave direction is not recorded, and hence not known; thus in determining the refraction coefficient, the wave direction is assumed to be the forecast direction. Recently, a wave-direction recorder has been designed on the principle of the Rayleigh disk and is now undergoing test. It is planned to install such instruments with the wave recorders in the near future, and thus eliminate this chance of error.

Since the shore wave recorders were established early in 1947 at Point Sur and Heceta Head to continuously record the character of the sea, records have been obtained which now permit a rigorous comparison of forecast and observed wave characteristics. Forecasting procedure was, in general, that given by "Wind Waves and Swell, Principles in Forecasting." Fetch, decay distance, and direction were determined by inspection of the isobaric pattern of meteorological maps, where fetch was taken as the distance over which the wind blew toward the forecast point with a range of 30° on either side of the true direction. Occasionally, the fetch was well defined by meteorological fronts, the rapid fanning out of the isobars, or a sharp curvature of the isobars, but usually the boundary was in question for as much as several hundred miles. In these cases, the condition giving a maximum fetch and minimum decay distance was taken. Wind direction was assumed to be at a 20° angle to the left of the isobars. The average wind speed was taken as six-tenths of the geostrophic wind, and was determined directly from the isobar spacing by a graphical overlay. Both wind direction and velocity were checked against ship reports in the area, and serious discrepancies corrected if possible. The wind duration was determined from previous weather maps and the forecast wave height in the fetch. Then, having the wind velocity and duration and the length of fetch, the wave characteristics at the end of the fetch were determined from the forecast graphs. It is interesting to note that, in the great majority of cases, wave generation was limited by the duration rather than by the length of the fetch. Having these wave characteristics and the decay distance, the decay graphs were entered, and the final wave characteristics determined. This method gives the characteristic, deep-water, wave lengths and periods. The direction from which the waves came was taken as the direction of the fetch from the forecast point, regardless of wind direction in the fetch.

The comparison at Point Sur covered April to December, 1947, and about 271 forecasts. At Heceta Head comparisons covered the period May to December, 1947, and involved about 201 forecasts (FIGURE 2). The actual comparison of recorded to forecast values was, in general, good, but still leaves something to be desired. A statistical analysis shows that 97 per cent of the recorded significant increases in wave height were forecast, but 23 per cent of the forecast wave trains failed to arrive. The rather large proportion of non-arrivals apparently resulted from the erroneous selection of fetches, frequently because of difficulty in ascribing limits to the effective

angles of the winds with respect to the recording point. However, to err in the direction of safety is somewhat gratifying. Sixty-nine per cent of the

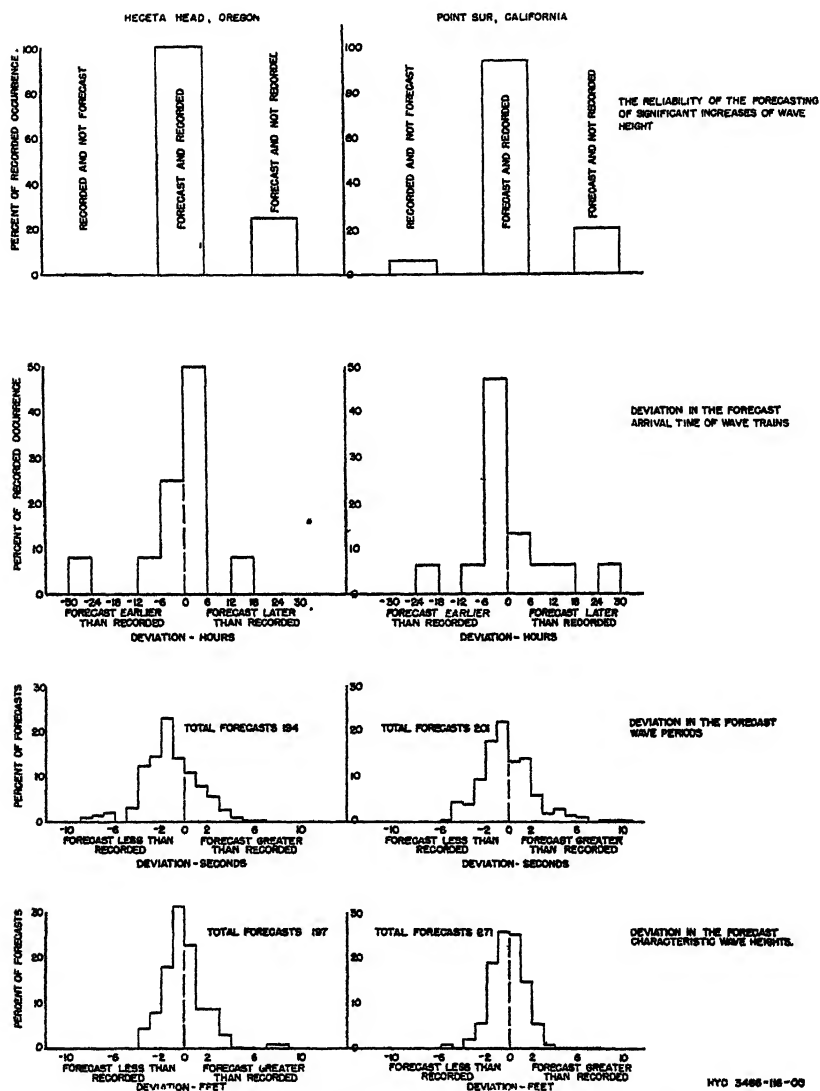


FIGURE 2. Comparison of forecast and recorded wave characteristics.

arrival times were predicted within 6 hours. The arrival times of the wave peaks were forecast within 6 hours for 39 per cent of the wave trains, and within 12 hours for 63 per cent.

The arrival times were generally predicted earlier than actually occurred—

especially those of the peaks of the wave trains, for which 64 per cent were forecast earlier, 24 per cent later, and 12 per cent were not forecast. The one exception was the predicted arrival of the wave trains at Heceta Head where 58 per cent were forecast later than recorded (of which, however, 50 per cent were within 6 hours of the recorded arrival time). Fifty-three per cent of the forecast wave heights were within 1 foot of the recorded heights, and 80 per cent within 2 feet. Sixty-three per cent of the forecast periods were within 2 seconds of the recorded periods. The forecast periods were usually lower than the recorded ones—at Heceta Head, 71 per cent were below the observed, and 59 per cent at Point Sur. In limited fetch work at Abbotts Lagoon, the periods from the Sverdrup-Munk curves also were consistently lower than the recorded values, and this may tend to show that the error lies in the constants employed in the equations of wave generation rather than in decay relationships or the application. The error in period decreases with waves of larger period. The deviation of predicted heights from the recorded values show no particular tendency to be higher or lower than recorded—the deviation curves being centered about the zero deviation line.

Few, if any, weather reports were made south of 30° latitude; none below 20° latitude. Hence, the isobaric pattern to the south was drawn by the dictates of probability and general experience rather than of actual data, and therefore is occasionally likely to be in error. Thus, it is probable that storms to the south, not observed on the weather map, generated some of the waves that were recorded, but not forecast (amounting to about 3 per cent of the total). This is substantiated by the arrival of wave trains in this category at Point Sur, but with none recorded and not forecast at Heceta Head, which is much less exposed to this southerly swell.

For several dates, weather maps have been obtained from the U. S. Weather Bureau and wave forecasts made from these as well as from the usual maps from the Navy Weather Central. These maps were drawn from the same data as the Navy maps, but analyzed by different weather forecasters. The forecasts from both sets of maps gave nearly the same results—but frequently bracketed the recorded data. This would seem to indicate that the smaller deviations of the forecast from observed values are, in a major part, due to the difference in analysis of the weather data, and the subsequent difference in the fetch and decay distances and wind velocities used in making the wave forecasts.

It may be of some interest to note the comparison of these recorded and forecast values with those obtained by The Scripps Institution of Oceanography in a statistical study of hindcast wave conditions along the California coast for three years. At Cape Blanco, Oregon (80 miles south of Heceta Head), Scripps found, for the summer months, that all the waves were under 8 feet, 96 per cent below 6 feet, 86 per cent below 4 feet, and 64 per cent below 2 feet. The forecast data for Heceta Head, Oregon, for the summer months

last year give 97 per cent below 8 feet, 96 per cent below 6 feet, 94 per cent below 4 feet, and 64 per cent below 2 feet. The recorded values for Heceta Head are 100 per cent below 8 feet, 98 per cent below 6 feet, 89 per cent below 4 feet, and 46 per cent below 2 feet. It is concluded, therefore, that existing forecasting technique results in a high degree of reliability for forecasting the arrival of significant increases in wave height, and prognosticating the wave heights. These two factors are the principal ones for the control of marine and shore activity. The forecast of wave period and arrival time of the peaks of the wave trains does not display the same degree of reliability evident in the former characteristics. This deficiency, or lack of perfection, does not detract from the exceptional value such forecasts possess to the people of the Northern Pacific coast, where daily activities, in a great many cases, are virtually dictated by what once could be known as the vagaries of the sea.

References

- ARTHUR, R. S. 1948. Forecasting Hawaiian swell from January 2 to 5, 1947. *Bull. Am. Meteorol. Soc.* 29(8): 395-400.
- BASCOM, W. 1947. Installation of a wave recorder at point Sur. Waves Project Report HE-116-242. Department of Engineering, University of California. (Unpublished)
- CORNISH, V. 1934. *Ocean Waves and Kindred Geophysical Phenomena*. Cambridge University Press. London.
- DEFAUT, A. 1929. *Dynamische Ozeanographie. Einfuhrung in der Geophysik*. Berlin.
- FOLSOM, R. G. 1946. Shore wave recorder drawings. Waves Project Report HE-116-219, Department of Engineering, University of California (Unpublished); 1947. Sub-surface pressures due to oscillatory waves. *Trans. Am. Geophys. Union*. 28: 6: 857-881.
- HERODOTUS. *The Persian Wars*, Bk. 7. Translated by George Rawlinson; *The Greek Historians*. Random House.
- IVERSEN, H. W. 1945. Sighting bar for surf observations of breaker height and distance to breaker. Waves Project Report HE-116-42. Department of Engineering, University of California. (Unpublished).
- JEFFREYS, HAROLD. 1925. On the formation of water waves by wind. *Royal Soc., Proc., A*, 107: 189-206.
- JOHNSON, J. W., M. P. O'BRIEN, & J. D. ISAACS. 1948. Graphical construction of wave refraction diagrams. Hydrographic Office.
- JOHNSON, J. W. 1948. The characteristics of wind waves on lakes and protected bays. *Trans. Am. Geophys. Union*. 29(5): 671-681.
- JOHNSON, J. W. 1948. *Proc. Am. Soc. Civil Engrs.* 74(8): 1401-09.
- KRUMHOLZ, OTTO. 1911. *Handbuch der Ozeanographie*. Stuttgart.
- LAMB, H. 1932. *Hydrodynamics*. (6th Ed.). Cambridge University Press. London.
- PATTON, R. S., & H. A. MARMER. 1932. *The Waves of the Sea; Physics of the Earth*, 5 Oceanography. National Research Council.
- ROSBY, C. G., & R. B. MONTGOMERY. 1935. The layer of frictional influence in wind and ocean currents. *Physical Oceanography* 3: 3.
- SCRIPPS INSTITUTION OF OCEANOGRAPHY. 1947. A statistical study of wave conditions at five open sea localities along the California coast. Wave Report No. 68.
- SVERDRUP, H. U., & W. H. MUNK. 1947. Wind, sea and swell: Theory of relations for forecasting. H. O. Pub. No. 601; 1947. Hawaiian swell from January 2-5, 1947; Scripps Institution of Oceanography. Wave Report No. 72.
- U. S. HYDROGRAPHIC OFFICE. 1944a, Wind, Waves and Swell: Principles in Forecasting; H. O. Misc. 11275; plates revised April 1947; 1944b. Breakers and Surf: Principles in Forecasting; H. O. Pub. No. 234.

VARIABILITY IN DIRECTION OF WAVE TRAVEL*

By ROBERT S. ARTHUR†

Scripps Institution of Oceanography, University of California, La Jolla, California

Introduction

A usual assumption in wave forecasting is that the waves in a fetch travel in a uniform direction coincident with that of the wind. Aerial photographs of wave trains (FIGURE 1) and visual and instrumental observations show a complex pattern of short-crested waves with waves of appreciable height traveling not only in the wind direction, but also in other directions.

In this paper, the assumption is made that, in a generating area, waves are present which are moving at various angles with the wind direction. The growth of these waves under the influence of the tangential stress and normal pressure of the wind is calculated by generalizing the relationship for wave growth developed by Sverdrup and Munk.¹

This variability in wave direction is not confined to the waves in the generating area, but is a feature which is retained after the waves emerge from the generating area and travel as swell. Since it is assumed that the direction of travel remains unchanged once the waves are formed, the variability in swell direction is determined by the variability in direction of the waves in the generating area.

When wave trains are interrupted by islands and headlands, such variability in the direction of wave travel is generally of greater importance in determining wave conditions in the sheltered region than refraction or diffraction. It is of critical importance in connection with longshore currents and problems of wave generation in small bodies of water of irregular shape.

Theory

Notation:

θ —angle between wind direction and direction of wave travel;

C —phase velocity of wave;

E —mean total energy of wave per unit area;

x —horizontal distance coordinate;

t —time;

R_T —mean rate of energy transfer to wave due to tangential wind stress;

R_N —mean rate of energy transfer to wave due to normal wind pressure;

U —wind velocity at about 8 m above sea surface;

ρ —density of water (1.025 g/cm^3);

* Contribution from the Scripps Institution of Oceanography of the University of California, New Series No. 408. This work represents results of research carried out for the Hydrographic Office, the Office of Naval Research, and the Bureau of Ships of the United States Navy Department under contract with the University of California.

† The writer is very much indebted to Mr. C. P. Ho for his assistance and suggestions in carrying out the numerical phases of this work.

ρ' —density of air (1.25×10^{-3} g/cm³);

γ^2 —resistance coefficient applicable to wind ($\gamma^2 = 2.6 \times 10^{-3}$ for $U' > 5$ m sec);

$\alpha = 2\gamma^2\rho' \rho$ (6.35×10^{-6});

g —acceleration of gravity (980 cm/sec²);

s —sheltering coefficient;

$\alpha = s/2\gamma^2$;

H —wave height (from crest to trough);

L —wave length;

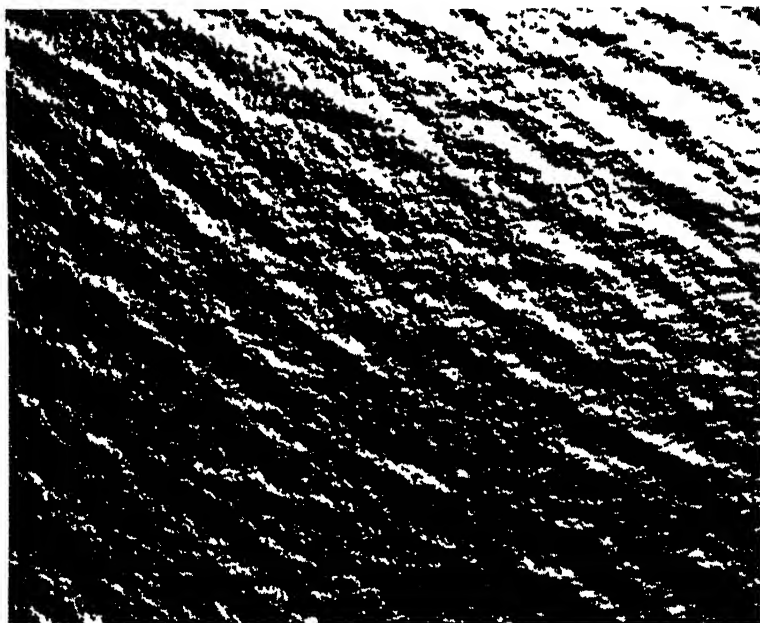


FIGURE 1. Aerial photograph of waves.

$\delta = H/L$ (wave steepness);

$\beta = C/U' \cos \theta$ (wave age);

r —coefficient of energy partition;

δ_1 —value of δ for $\beta = 1$;

β' —a value of β determined by the assumed δ -value for $\beta = 0$;

F —length of fetch in direction of wave travel.

The Energy Equations and Their Solutions. Assume for the equations expressing the energy budget of a train of waves, following the method of Sverdrup and Munk:

Steady state:

$$\left. \begin{aligned} \frac{C}{2} \frac{dE}{dx} + \frac{E}{2} \frac{dC}{dx} &= R_T \pm R_Y \end{aligned} \right\} \begin{aligned} &+ C < U' \\ &- C > U' \end{aligned} \quad (1)$$

Transient state:

$$\left. \begin{aligned} \frac{dE}{dt} + \frac{E}{C} \frac{dC}{dt} = R_T \pm R_N \end{aligned} \right\} \begin{aligned} + C < U \\ - C > U \end{aligned} \quad (2)$$

Assume

$$R_T = EA g U^{-1} \beta^{-3} \sec^2 \theta; \quad R_N = EA g U^{-1} \beta^{-3} \alpha (1 - \beta)^2 \sec \theta \quad (3)$$

where, in the case of the tangential stress, the stress component in the direction of wave advance is introduced, and, in the case of the normal pressure, the wind component in the direction of wave advance is introduced.

The solutions of the energy equations are obtained in exactly the same manner as that used by Sverdrup and Munk, and the details will not be repeated here. The authors just cited assume that wave steepness, δ , is a function of wave age, β , only, and proceed to evaluate certain constants which occur in a theoretically derived relationship by fitting observational data.

The further assumption is now made that approximately the same numerical relationship between δ and β holds for values of θ other than $\theta = 0^\circ$. This requires redetermination of the constants r , α , δ_1 , and β' for each value of θ . The values of the constants for $\theta = 30^\circ$ and $\theta = 45^\circ$ as compared to those obtained by Sverdrup and Munk for $\theta = 0^\circ$ are as follows:

θ	r	α	δ_1	β'
0°	0.580	2.50	0.0380	0.350
30°	0.563	3.14	0.0393	0.344
45°	0.544	4.19	0.0402	0.339

The solutions can all be presented as relations between non-dimensional ratios, and the results in such form for $\theta = 30^\circ$ and $\theta = 45^\circ$ are compared to those of Sverdrup and Munk for $\theta = 0^\circ$. FIGURE 2 shows the parameters C/U and gH/U^2 plotted against gF/U^2 (F is the fetch in the direction θ), that is, wave speed and height as functions of fetch and wind velocity assuming unlimited duration. FIGURE 3 shows in terms of the non-dimensional ratios C/U , gH/U^2 , and gt/U , the wave speed and height as functions of duration and wind speed, assuming unlimited fetch.

Discussion

In determining the height and period of waves traveling in a direction other than the wind direction, it is convenient to make use of the ratios H_θ/H and T_θ/T , where H_θ and T_θ are the height and period respectively of the waves traveling over a distance F at an angle θ to the wind direction, and H and T are the corresponding height and period of the waves traveling *with* the wind over the same distance. The distance F is measured in the direction θ . An alignment chart (FIGURE 4) gives H_θ/H and T_θ/T as functions of wind velocity U , and either fetch F (steady state) or duration time t (transient state) for θ equal to 30° and 45° . An example is shown.

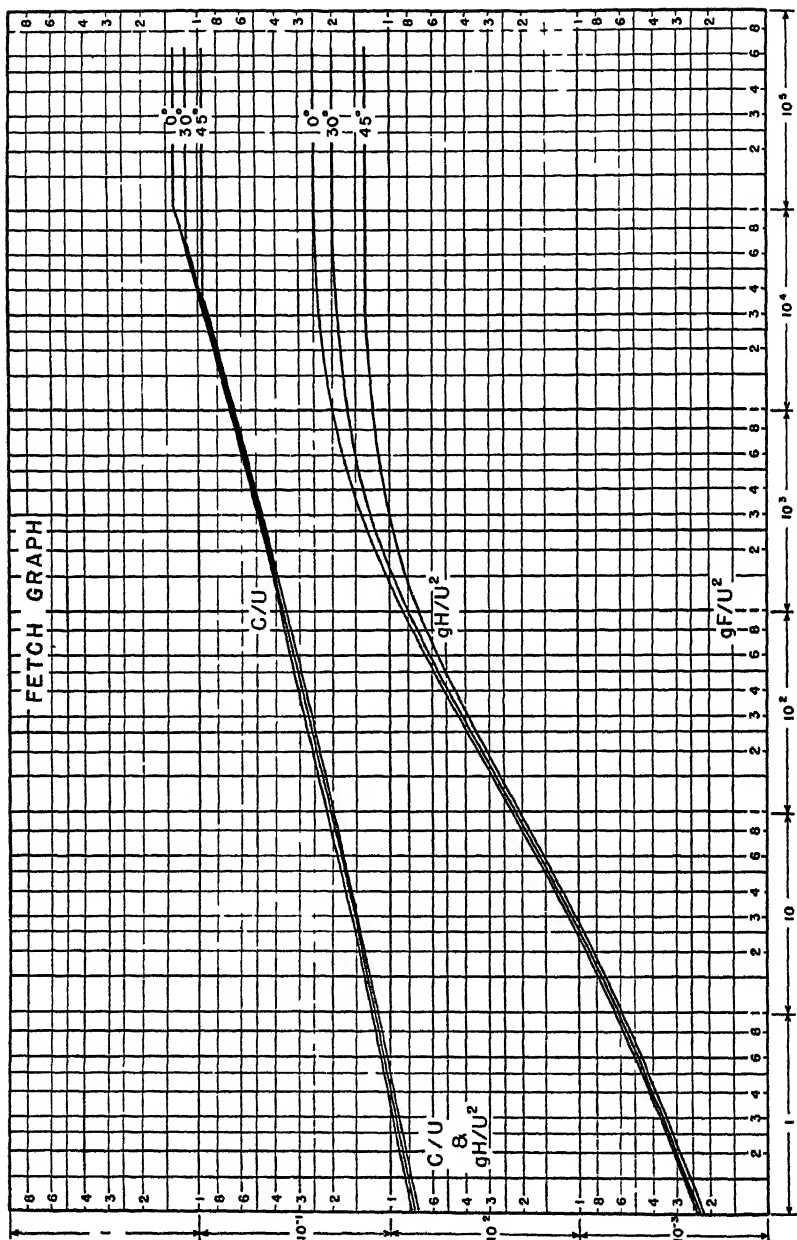


FIGURE 2. Curves giving wave velocity and wave height as functions of fetch and wind velocity in terms of non dimensional parameters for angles of 0° (after Sverdrup and Munk), 30° , and 45° .

In order to determine H_θ and T_θ , it is first necessary to find H and T by means of Plates I and II of Scripps Institution of Oceanography Wave Report No. 73,² or from the non-dimensional representations for $\theta = 0_\theta$. The question as to the existence of a steady state or transient state in the fetch is settled at the same time. H_θ/H and T_θ/T are then determined from FIGURE 4 and a simple multiplication gives H_θ and T_θ . The modification of H_θ and T_θ as the waves decay through a certain distance D is determined by means of the decay diagram^{1, 2}.

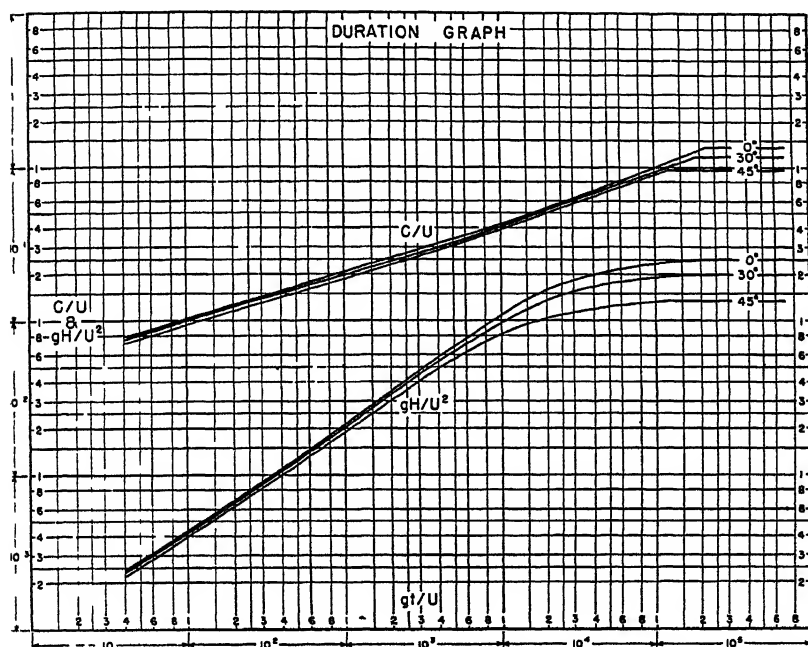


FIGURE 3. Curves giving wave velocity and wave height as functions of duration time and wind velocity in terms of non-dimensional parameters for angles of 0° (after Sverdrup and Munk), 30° and 45°

FIGURE 4 shows that as the angle θ increases, the wave height and period decrease. The percentage decrease is less for higher wind velocities, and is less for period than for height. The waves moving at an angle of 30° with the wind have heights between 80 per cent and 90 per cent of those moving with the wind for typical values of the forecasting parameters. For 45° , the heights are between 55 per cent and 75 per cent. Since waves of different heights are present in the fetch, it is to be expected that the higher waves which travel more nearly in the direction of the wind shelter the lower waves. FIGURE 4 therefore may be expected to give values of H_θ somewhat too large for small values of the ratio H_θ/H . No adequate observations are available from which a quantitative estimate of the effect may be made. Relationships for θ larger than 45° , have not been reproduced for this reason.

The present analysis assumes that, in a fetch, waves are moving in every direction which does not deviate from the wind direction by more than 45° . Since the wind can transfer only a finite amount of energy to wave

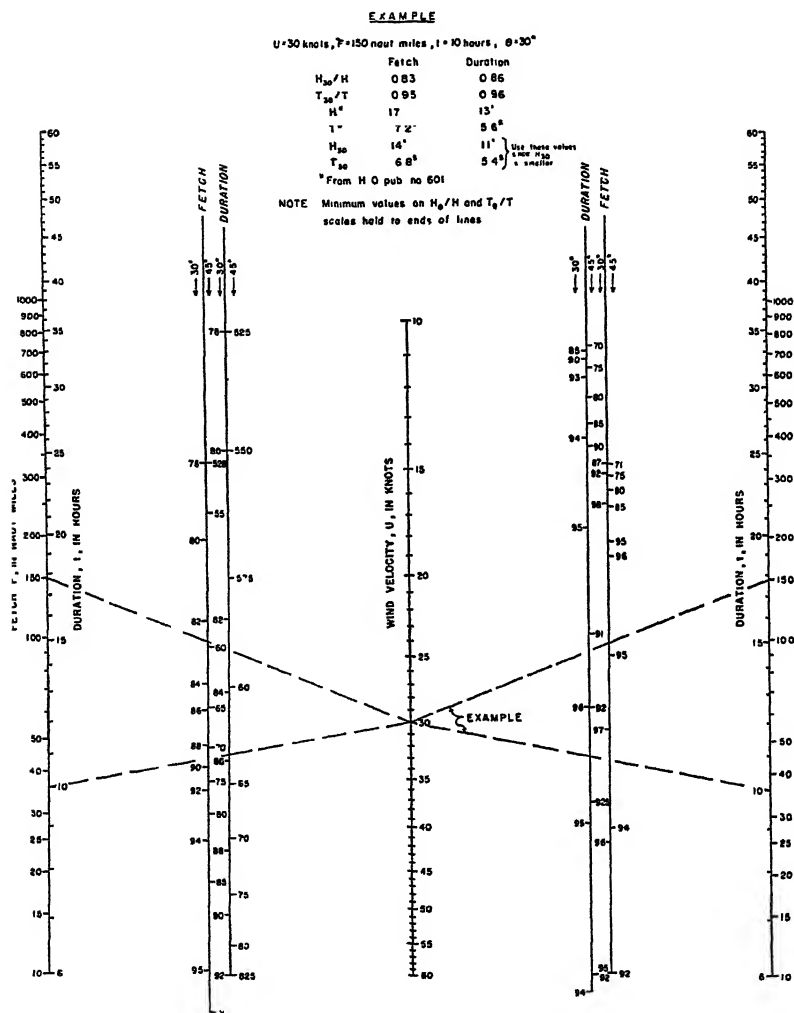


FIGURE 4. Alignment chart for the computation of height and period of waves moving at angles of 30° and 45° with the wind.

growth, the waves moving in a given direction must show height and period variability with time. The quantities H_θ and T_θ may be expected to apply to waves which correspond to the "significant waves" discussed by Sverdrup and Munk.¹ It is to be noted that the frequency distribution of the wave height and period spectra may change with the angle θ .

Application

Comparison with Forecasting Experience. Forecasting experience³ has demonstrated the need for considering that waves decaying from a fetch where wind direction is uniform reach not only points in line with the wind (A of FIGURE 5), but also those at angles up to 30° (B of FIGURE 5). The same height and period values have been used regardless of angle. This

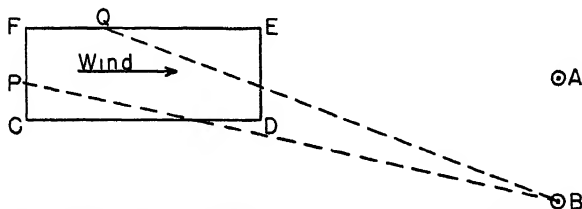


FIGURE 5 Schematic representation of a fetch area CDEF from which waves arrive at points A and B.

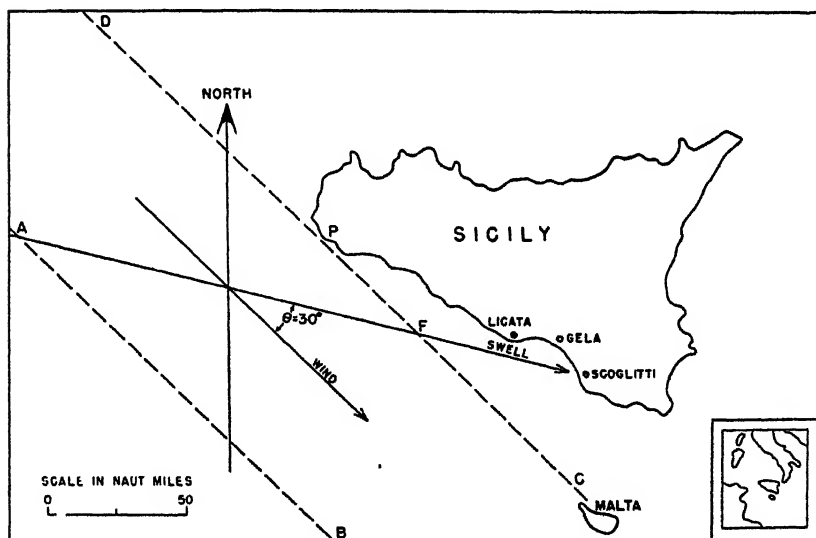


FIGURE 6. Origin of swell at the Scoglitti beachhead during the invasion of Sicily.

practice is in rough agreement with the variability results discussed above, since, for θ less 30° , the value of H_θ/H is generally greater than 0.80, and T_θ/T is greater than 0.95.

Invasion of Sicily. Data from a report on the invasion of Sicily⁴ can be used to make a check on the theory. The weather maps show that any swell affecting the landing near Scoglitti must have been generated in region ABCD (FIGURE 6) by the north-northwest to northwest winds whose velocities were as high as 31 knots up to 2000 GCT, D-1 day. The winds in

the region to the east of line CD and south of Sicily were also from the northwest but velocities were less than 16 knots, because of the sheltering effect of the island. The only swell which could approach the beach at Scoglitti would be that refracted or diffracted from point P, at the southwest corner of the island, or that proceeding from region ABCD as a result of variability in direction of travel. Because of the bottom topography and the distance involved, the "refraction factor"⁵ associated with waves refracted from P to Scoglitti would be very small. Diffraction studies show that the "diffraction factor"⁶ would also be very small. However, relatively high swell would be expected off Scoglitti as a result of variability in direction of travel of waves generated in area ABCD. The highest swell to be expected would be generated over a path AF with θ equal to 30° . The computation of height for point F of waves moving at an angle of 30° to the wind is as follows:

F	U	t^*	H	θ	H_0/H	H_θ
150 miles†	30 knots	10 ^h	13'	30°	0.86	11'

A following wind is present as the waves decay a distance of about 50 miles from point F to Scoglitti. It is assumed that the following wind is sufficient to keep the swell from decreasing in height over the decay distance. As a result, the expected height of the swell off the beach at Scoglitti would be 11 feet. The travel time would be about 5 hours so that the swell would arrive at approximately H-hour. Observations during the early landings showed a swell height of 10 to 12 feet.

San Clemente Island. Observations of wind waves were made around the southern end of San Clemente Island, California, on June 30, 1947 from the E. W. SCRIPPS, research vessel of the Scripps Institution. Stations A and B (FIGURE 7) were occupied at 0618 and 0730 PST respectively. Other stations to the east of the island were also occupied to determine the island's effect on westerly swell, but no swell was discernible on this occasion. The wave data (FIGURE 7) were obtained by visual observation of a short-line-spar buoy wave meter.

Station A is exposed to waves generated in the fetch to the northwest by the winds which at this season are predominantly northwesterly. This fetch is interrupted by San Nicholas Island, 50 miles to the northwest, and the Santa Barbara Islands, 100 miles to the northwest. However, a small change of angle gives an unobstructed fetch. Assuming steady state conditions, the following values of the forecasting parameters are consistent with the observed height and period at Station A:

Wind Direction	U	F	Wave Direction	θ
NW	14 knots	250 miles	NW	0°

* The actual duration time of the high wind velocities has been extended to allow for the waves already present in the fetch.⁷

† All miles referred to are nautical miles.

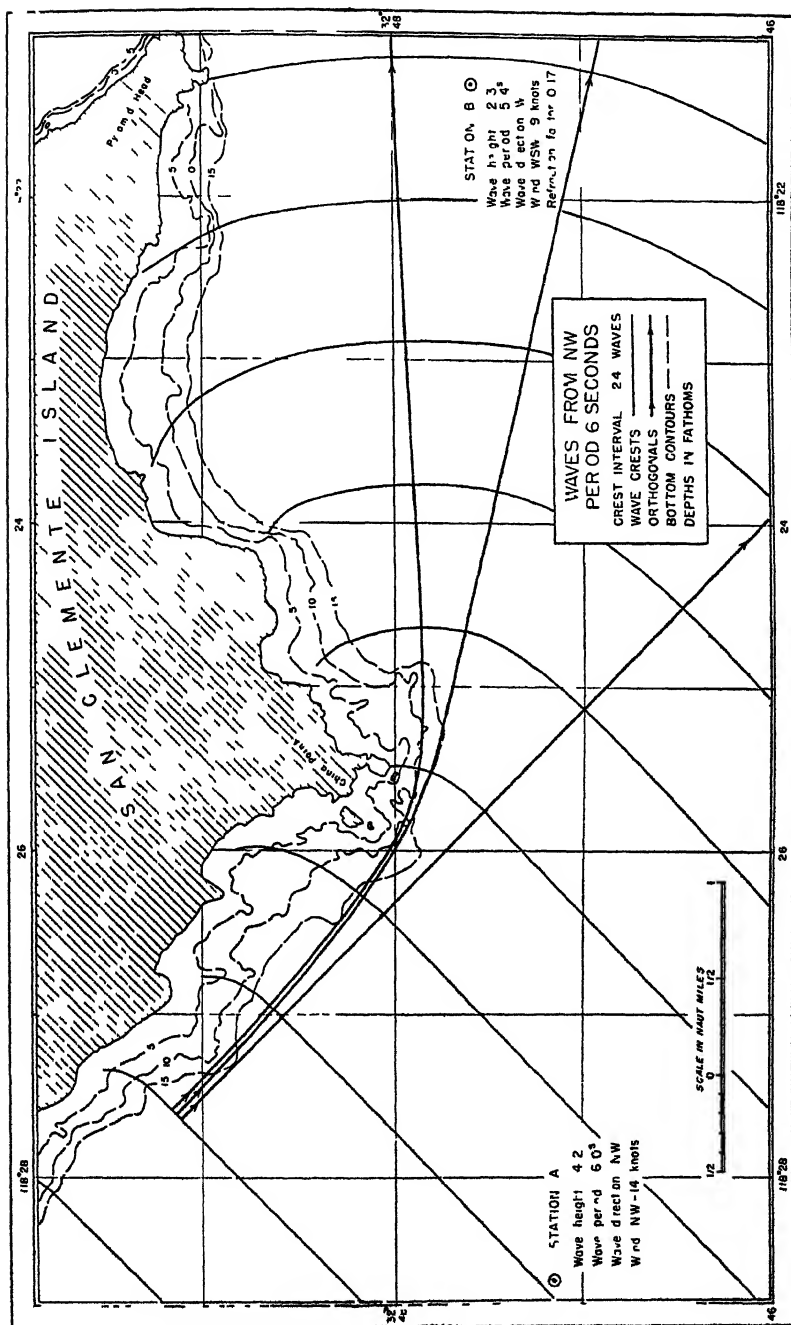


Figure 7 Refraction diagram for China Point region, San Clemente Island, with wind wave observations for both a sheltered and unsheltered station

Because of China Point, waves can reach Station B only from due west or southerly directions. In view of the wind field, the waves observed at Station B must be a result of refraction or diffraction at China Point, or variability in direction in the northwesterly fetch. Since the refraction coefficient is 0.17 and the diffraction coefficient is less than 0.10, these processes cannot account for the observed waves at B. The maximum

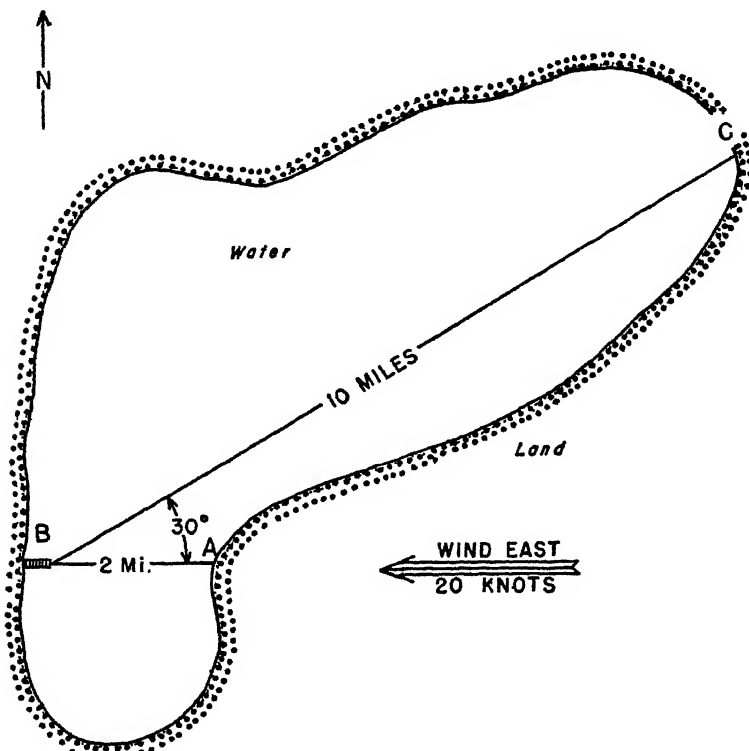


FIGURE 8. Schematic diagram illustrating the "direct" fetch, AB, and the "effective" fetch, CB, for waves reaching point B in a small body of water of irregular shape.

height as a result of variability in direction would be associated with waves from the west. The computation of these waves from the above forecasting parameters is as follows:

θ	H_0/H	T_0/T	H_0	T_0
45°	0.55	0.76	$0.55 \times 4.2 = 2.3'$	$0.76 \times 6.0 = 4.6''$

The values compare favorably with the observed height of 2.3 feet and period of 5.4 seconds.

Conclusions

The present results indicate that, within the fetch, waves may be expected to grow, even though they are moving at angles of as much as 45° with the

wind direction. The height of such waves will be at least 50 per cent of the height of waves moving with the wind. Within the fetch, the waves moving with the wind or at small angles to the wind appear dominant, but waves moving at larger angles are important when a wave train is interrupted by an island or headland. Variability in direction of swell can be determined by considering the different paths in the fetch which are directed toward the point where swell is to be considered. Lines PB and QB in FIGURE 5 represent two such paths. Variability in direction of swell may be expected to decrease with increasing decay distance.

FIGURE 8 illustrates schematically a critical case in connection with variability in a body of water of irregular shape. Assuming steady state conditions, the waves reaching point B from along the "direct" fetch AB have computed heights of 1.5 feet, whereas those reaching B from along the "effective fetch" BC have computed heights of 3.5 feet. No checks have been made on the application of the principles given here to small bodies of water, and it is recommended that the results be used only as a guide until such checks have been made.

Bibliography

1. SVFRDRUP, H. U., & W. H. MUNK. 1947. Wind, sea and swell; theory of relations for forecasting. U. S. Hydrographic Office, U. S. Navy, Tech. Report 1. H. O. Pub. 601.
2. ARTHUR, R. S. 1947. Revised wave forecasting graphs and procedure. Scripps Institution of Oceanography of the University of California. Wave Report 73.
3. HYDROGRAPHIC OFFICE, U. S. NAVY. 1943. Wind waves and swell—principles in forecasting. H. O. Misc. 11, 275.
4. CHIEF OF NAVAL OPERATIONS, AEROLOGY SECTION. 1944. The invasion of Sicily. NAVAER 50-30T-1.
5. MUNK, W. H., & M. A. TAYLOR. 1947. Refraction of ocean waves: a process linking underwater topography to beach erosion. J. Geol. LV(1): 1-26.
6. PUTNAM, J. A., & R. S. ARTHUR. 1948. Diffraction of water waves by breakwaters. Trans. Am. Geophys. Union. 29(4): 481-490.

Discussion of Paper

DR. W. L. DONN (*Brooklyn College, Brooklyn, N. Y.*): In the course of studying wave and swell records for the Woods Hole Oceanographic Institution, with the purpose of correlating swell observations with marine weather conditions off the eastern coast of the United States, certain inexplicable discrepancies were noted which appear explainable in the light of the wave variability theory.

The records were made by a bottom-resting, pressure-type wave meter, about a mile southeast of Cuttyhunk Island, near the western end of Cape Cod. The meter was constructed and operated by the Woods Hole Oceanographic Institution. The weather correlation was attempted on the basis of the North Atlantic Weather Maps prepared by the U. S. Weather Bureau at La Guardia Field, Long Island.

The exposure of the wave meter was such that the only "window" for the free approach of waves from a distant source was from a sector bounded by lines extending to the southeast and southwest. A fetch of but a few miles of shoal water exists to the north and northwest of the recording

station. To the west and west-southwest, in the directions of Long Island Sound and Long Island respectively, a possible fetch of more than 100 miles occurs, with water depths averaging close to 20 fathoms.

Most of the significant swell recorded on the Cuttyhunk meter (periods from 6 to 10 seconds) could be correlated with synoptic situations in which cyclonic warm sectors existed off or along the eastern coast. In such cases, the southerly or southwesterly winds which prevailed had a relatively long fetch along the coastal waters. The duration of these winds was normally closed with the passage of a cold front. The strong northwesterly winds which followed the fronts rapidly damped the pre-existing southerly swell.

In certain cases, with strong west to west-northwest winds following the frontal passage, the wave records either did not show the expected damping, or showed new waves developing of the same order as those previously existing. This was difficult to explain in view of the up-wind fetch being limited to a few miles by the coastline of Connecticut. However, if a deviation of wave propagation from wind direction of 30 degrees is permitted, the effect of the strong winds over the longer fetch to the west and west-southwest, mentioned above, might have permitted the development of the observed waves.

These observations appear explainable from qualitative considerations, although more complete quantitative studies will be made to check the observations further.

OCEAN WAVE RESEARCH AND ITS ENGINEERING APPLICATIONS

By M. A. MASON

Beach Erosion Board, Army Engineers, Washington, D. C.

Ocean waves are one of the natural phenomena of our environment that have attracted the attention of scientists and laymen almost from the time of our first recorded history. Yet, curiously, until the last few years, little progress, and in fact, limited study, leading to explanation of ocean wave phenomena have been made. Such knowledge as we had until recently was largely the result of studies by mathematicians and physicists, stemming from their curiosity about natural phenomena aided and abetted by various other interests. Again, curiously, a large part of such knowledge was based on reasoning rather than observation. Perhaps this situation can be attributed to the nature of the phenomena to be observed, for as Harold Jeffreys, one of the modern contributors to our pre-war knowledge of ocean waves, once wrote after observing the sea, "The predominant characteristic of the sea is its irregularity." Those of us who have been trying to observe and define the character of the sea surface have no quarrel with Jeffreys on this point.

This paper will be an attempt to consider the relations between our present knowledge of ocean surface waves as developed by research, and the engineering applications of that knowledge. Some attention will be given to deficiencies or gaps in our knowledge of waves which are needed for the solution of various engineering problems.

I shall not attempt to review the meagre fund of our pre-war knowledge. Those who work in the field know it well. Let it suffice simply to remind you that, so far as engineering applications were concerned, the knowledge we had was rudimentary almost to the extreme.

We might profitably consider for a moment the engineering applications we have in mind. Oldest in point of view of time is the application of wave knowledge to the design and operation of ships. It is surprising to one who is not a naval architect to discover how slight a foundation of knowledge of natural wave characteristics underlies the design of ships. So far as I can discover, there seems to be available among ship designers some fair knowledge of the probable maximum wave lengths, periods, and heights ordinarily encountered in the various seas, but probably not much other detail. There also have been developed by experience certain rules of operation of vessels to minimize passenger discomfort and danger of vessel damage, and to increase the operating economy of the ship, but full possibilities do not seem to have been realized.

Perhaps the next oldest problem is the application of knowledge of ocean disturbances to the location and layout of harbors, and in the design of

harbor protective works. This problem has been most vexing, as is best attested by the difficulties associated with the operation and maintenance of some coastal harbors, both in this country and abroad. The Permanent International Association of Navigation Congresses has given prominent, if not primary attention, to problems of this sort for many years, particularly to the design of protective works after the disastrous experiences at Bizerte and Antofagasta. The frequently tragic consequences of protective works failures give breakwater design problems a great public interest, but the problems of harbor location and layout with respect especially to wave action to insure safe, sheltered harbors at lowest cost are equally important.

From the point of view of natural resources, the most important engineering application of ocean wave research lies in the field of shore protection, including in the term not only protection of the actual shore against encroachment by the sea, but also the provision and maintenance of navigable entrances to rivers, bays, harbors, etc. No one knows what the total cost of inadequate shore protection may be, certainly it is very large in this country. In the Netherlands, it may be said to be essentially the capital value of a large portion of the country. Great Britain has, for some years, been alarmed at the inroads of the sea, having formed their Royal Commission on Coastal Preservation to study the situation, and spent millions of dollars to bulwark their rapidly eroding land.

The military services are concerned with a host of engineering applications of ocean wave research to their own highly specialized problems. In this connection, it is interesting to note the extreme disrepute of oceanographic knowledge in the military services generally, prior to World War II, and the relatively favored position it enjoys today. Some of the problems interesting the military services are most unusual. One which I have in mind is the effect of sea surface disturbances on radar operation—a subject recently reported upon in the English literature. Other problems are more familiar to us, for example: the effect of surf on the casualty rate of landing craft in amphibious operations; the importance of wave and surf knowledge in the selection of landing beaches; the application of wave knowledge to airplane carrier operation; and many others.

We might continue these illustrations, but, since I wish to discuss some of them in more detail later, let us, rather, review our state of knowledge as developed by ocean wave research.

Our chief objective remains to be reached. In spite of familiarity with the problem for several thousand years, we still find ourselves in the indefensible position of attempting to study a natural phenomenon which we have not adequately defined. Borrowing from Thorade, we may repeat in 1948 what he wrote in 1931—"No adequate results of observation are available with respect to form, orbital path, and energy, but they are sufficient to shake our confidence in theory . . . Complete agreement between theory and observation is seldom found, and where it is found, it seems suspicious."

What is an ocean surface wave? I believe I can truthfully say that we cannot define an ocean surface wave from observation. You and I have watched waves on the Pacific Coast that seemed to be regular and, to some extent, uniform, but even casual observation showed that consecutive waves varied widely in period, form, and estimated height (an observation confirmed by many instrumental measurements of waves on the same coast).

As an engineer concerned with problems in which ocean waves are important, I am not primarily interested in an individual ocean wave, or even single systems of waves. However, I must know in detail the characteristics of changes in sea surface elevation and associated internal movements resulting from wave action and the temporal and spatial variations of these characteristics. Therefore, I am vitally interested in and must support observations, not primarily of ocean waves, but of the changes in sea surface elevation associated with ocean wave action; *i.e.*, the definition of the characteristic irregularity of the ocean surface.

As positive knowledge, we know many more things than we did some years ago. We can predict, with an accuracy that is often sufficiently good, the occurrence and probable time of arrival of certain waves resulting from a known meteorological pattern. We can forecast the probable character of the surf those waves will form when they reach the coast. Refraction diagrams, in simple cases, and sometimes diffraction diagrams can be drawn to indicate the approximate distribution of the available wave energy over some area of shoal water or coast. If sufficiently good instruments are employed in certain ways, we believe we can detect the so-called forerunners of swell and from them predict the location of an atmospheric disturbance. These and a few other things we know how to do, but most importantly, we think we can measure ocean waves.

Let me dwell on this achievement a moment. We can, by surface or sub-surface instruments, obtain a measure of the variation with time of the elevation of the sea surface at a point. If we attempt to analyze this measurement record, however, we are in a dilemma. For we must adopt the subterfuge of concerning ourselves only with the highest and longest waves, or I should say, more rigorously, the most prominent changes in sea surface elevation, in order to arrive at numbers to define the waves. The concept of significant waves developed by Sverdrup, for example, is most useful for some problems, but it does not relieve us of the necessity to seek a more reliable and accurate means of identifying waves. An analysis based on the time rate of variation of energy passing our measurement point would be more useful in most engineering applications, but greatly more difficult to effect. It must be noted that, when we claim to measure waves by recording surface elevations at a point, we require that the waves do not change shape or velocity as they pass the point, and that the various wave systems which may co-exist and interact can be separated through analysis of the record.

Neither of these conditions is satisfied for shallow water waves, and the latter is not satisfied for deep water waves, although the former may be.

Suppose we return to the engineering applications of ocean wave research and see where we stand. Entirely aside from our technical interest, which we as scientists believe to be paramount, I believe such an examination to be important in allowing us to evaluate the support to be given to continued ocean wave research. Perhaps the injection of this realistic note into our discussions is to be decried. However, our present state of relative affluence in research funds already shows some signs of strain; and we must realize that not for long can we continue projects, unless they can be defended against competing projects, in this and other fields of scientific inquiry. Engineering applications, either immediate or reasonably foreseeable, are very good arguments for support of projects in terms of funds and brain-power.

Let us consider, then, what value knowledge derived from ocean wave research might have in engineering applications, and examine, first, its value to the designer and operator of surface ships. A naval architect has the problem of designing a hollow, self-propelled body to carry a specified load and to operate at the interface of two fluids, water and air. Certain requirements limit the water submerged portion of the body to a depth of 35 feet or so. This means that the submerged portion of the ship is always operating in a region where the effects of surface disturbances or waves are most pronounced. The designer, before constructing the vessel, subjects models of hull designs to towing tests to determine its still water "skin-friction" resistance and its "wave-making" resistance. The behavior of the hull in artificially generated uniform trains of waves is sometimes observed. It is known, of course, that these waves do not represent ocean surface conditions. But what are the typical conditions for the Atlantic, Pacific, Indian Oceans and other seas? Can they be reproduced in the model basin and, if so, how? Is it worthwhile to reproduce them? Knowledge of ocean wave conditions in the open sea will enable the study of these questions among others. The steering rudder and the propeller operate in the region of maximum surface disturbance. What are the forces commonly exerted on rudders by wave action? How does wave action affect propeller efficiency and what can be done to modify these effects? To start to study these questions we must have knowledge not only of the true surface conditions but, in addition, quantitative information on the velocity fields, to at least the 35-foot depth, associated with the surface wave conditions.

The designer must build his ship sufficiently strong to carry its appointed load when subjected to the most violent conditions of the open sea. The frequency of failures of war-built ships illustrates the importance of this consideration. If the designer could be furnished information on the character, magnitude, and time rate of variations in surface elevation associated with maximum wave action, he would be in a more favorable posi-

tion than he is now to calculate the stresses and strains of racking, pitching, and rolling actions.

Captain Harold Saunders, until recently Technical Director of the David Taylor Model Basin of the U. S. Navy stated in a recent paper that "The most frightening lesson learned from the war just concluded, . . . was the almost desperate lack of basic information on the fundamentals of the nature of the sea and the behavior of bodies in it." He continues, listing some specific problems in hydromechanics of immediate interest to those concerned with ships on the sea, among which are:

The formation and behavior of spray, as affecting the operation, preservation, and habitability of objects in and around the sea.

The phenomena and kinetics of wave slap, wave impact, and breaking waves.

The elimination of surface waves in water streams.

The control of systems of ship-roll stabilization.

Obviously the solution of these engineering problems involves, in the first instance, ocean wave research and, as an important primary phase of the research, the definition of surface conditions. Much more might be said of applications of wave research in marine engineering, particularly of the possibilities of improving ship operation, but such discussion must be deferred to more appropriate circumstances.

There was mentioned earlier the close relation between the location and layout of harbors and ocean surface disturbances. This relation is illustrated by the case of Long Beach Harbor, California. The harbor is formed by a rubble mound breakwater running offshore from Point Fermin and partially enclosing the head of a spiral shape bay. During the war, it was noted that ships berthed at a certain dock were being damaged by reason of apparent surge movements in the harbor. Study of the problem led to the conclusion that the movements were caused by a swell from offshore of almost imperceptible height and with a period approximating 15 minutes. Other swell of about 5 minutes period was believed to be present and of some importance.

This same harbor was the locale of violent surge action, some years ago, when a very low, very long swell, which was not detected offshore, passed through the porous, rubble mound, offshore breakwater and severely damaged several vessels berthed in the harbor. It may be noted that the southern California shore generally is subject from time to time to savage attacks by long period, low amplitude waves apparently coming from the South Pacific.

Little is known about these very long period waves, even less is known of how they are generated. An interesting conjecture is the possibility that surface waves, under conditions of very long fetch, in large basins, where multiple reflections are possible, may eventually combine to give the appearance and effect of long waves. It is, of course, equally possible that

there is a complete spectrum of wave periods present, extending from ripples, through surface waves, to long waves. Study of the existence and nature of these unusual waves on the southern California coast would decidedly be of interest in allowing proper defensive measures to be taken.

But how could wave research have improved the Long Beach harbor surge situation? Let us assume that we had knowledge of the offshore wave conditions while planning the harbor. This information, with knowledge of the hydrography of San Pedro Bay, would have allowed the preparation of wave refraction diagrams showing the distribution of wave energy at the shore in the natural state. Similar diagrams prepared for the conditions of the complete harbor, as planned, would have emphasized any locations of concentration of energy. The layout of berths could then have avoided these locations, or the breakwater plan could have been modified.

Wave research can be of value now. It can define for the engineer planning expansion of the harbor, the nature of the wave action in the area, and enable investigation of the effects of this wave action on the harbor area for various locations and sizes of navigation entrances through a considered extension of the offshore breakwater to form a totally enclosed harbor. Research into the wave-generating possibilities within the harbor, combined with wave action through entrances, might result in ideas for solving the problem of sanitary flushing, or water exchange, to prevent pollution of the harbor area.

Study of surface and long wave propagation through porous breakwaters would be of great value to the Los Angeles-Long Beach Harbor problem, as well as many others. In fact, the whole subject of wave action on harbor protective structures is one of prime importance to engineers and harbor designers. The problem is one of great economic, as well as technical, interest, since breakwaters usually cost hundreds of thousands of dollars. Even small improvements in design, and sometimes a simple change of specifications, can well pay for the cost of the necessary research. In designing a breakwater, the usual problems are: where to locate the structure; what material to use for its construction; what cross section shape to give the structure; to what height shall the structure be built; and what is the maintenance required to keep the works in serviceable condition to perform their function. Each of these problems requires knowledge of the surface waves at the contemplated location. Since the primary function of the structure is to protect a given, and usually arbitrarily selected, area from wave action, every feature of the works must contribute to this purpose.

I have mentioned the Los Angeles-Long Beach Harbor case. It is not unique. The Great Lakes area has Grand Marais Harbor and others. The Atlantic Coast, though well provided with natural harbors, has its problems of maintenance of protective works. Hilo Harbor in Hawaii has a surge problem, and many others are subject to various types of problems arising from wave action.

In this field, there are many problems, too many to discuss them in detail, but a few are so fundamental and their solution, at least in part, is so badly needed that some discussion is warranted.

There has been indicated a relation between wave action and harbor location. The relation usually does not control selection of a location, because of the greater influence of navigation, transportation, or other benefits associated with a certain location. However, a thorough study of the site in relation to its wave conditions is conceded to be a requisite, though we engineers are sure that we are not exhausting the possibilities of this type of analysis. We would like to be able to know the direction of approach, period, and height of the probable maximum storm wave action; whether or not the waves would appear in uniform or variable series. Will there be several wave trains from differing directions co-existing? If so, will they inter-act to result in very high waves or breakers in the entrance? Suppose we must use quarried stone for our structure. How big must the stone be to withstand wave action? Can it be simply dumped, or must it be placed? Should it be quarried as a rough cube or some other shape? A problem to breakwater designers, the solution of which they are ignorant, is that of what slope should be given the seaward face of a mound breakwater. The saving in stone by increasing the slope from 1 on 2 to 1 on $1\frac{1}{2}$ would sometimes pay for the research required to answer all the questions posed.

Admittedly, these are engineering problems, questions of design rather than of the basic knowledge the research scientist seeks, but they are also problems whose solution is almost wholly contingent upon basic knowledge of surface wave phenomena. Stones weighing from 15 to 20 tons have been transported bodily on the Arecibo, Puerto Rico breakwater by swells arriving from the North-East, without a cloud in the sky at the site. The knowledge of where those swells came from, how often they may be expected, what the characteristics were which made them destructive, all this and more we can rightfully expect to be developed by research on ocean waves, the cost for research being only a minor fraction of the resulting benefits.

It is in the process of attempting to prevent encroachment by the sea on a coastline that we are most concerned with ocean or lake wave action, for it is the wave action which is primarily responsible for continual degradation of our shores, with its attendant loss of property and life. Huge waves batter at seawalls, hundreds perish as hurricane waves sweep over Galveston, the Florida Keys, or wherever. These are familiar headlines. Less familiar and less spectacular are the observations of shore students: shore line retreating 20 feet annually in Ohio; the New Jersey coast receding an average 2 feet each year; newly constructed seawalls damaged; or jetties trapping sand at a rate of 250,000 cubic yards annually. Waves do all these things, at a net cost in land per year of 1 foot along each of our own 52,000 miles of shore line, and a dollar cost in the millions.

The shore problem, if we can think of it in general terms, may be stated

in a simple fashion. There are two major elements in the problem: one, the material, sand, clay, rock, constituting the shore; and two, the energy available to modify and transport the material from one place to another, this energy being overwhelmingly supplied by the wave action at the shore. A further breakdown leads to the following six questions, which I formulated about a year and a half ago as a blueprint for the analysis of shore problems. If we can answer these six questions we can solve our problem. They are: (1) What are the sources and character of the beach material? (2) What are the rates of supply and loss of material to and from the shore? (3) What are the mechanics and manner of movement of material from the source to the shore and from the shore to other areas? (4) What are the feasible methods of modifying the rates of supply and loss of material to achieve the desired results, and what are the effects of such modification? (5) What are the design requirements of the feasible methods of modifying supply and loss rates? (6) What is the economic cost of each of the feasible methods of modifying rates of supply and loss?

One of the most powerful methods useful in this analysis is to formulate what might be called the material-energy balance of the area in terms of time. To do this, we must know the character and quantity of available material. This material is moved chiefly by the application of energy in the form of wave action. We purposely omit oceanic, tidal, or other currents in this general discussion, since it is believed they are important only in certain restricted cases. Little is known of the mechanics and manner of material transportation by wave action, and perhaps still less is known of the processes of derivation of material through wave action. In spite of this situation, much has been done, in the way of modifying rates of supply and loss, by modifying the energy balance. For example, considerable success has been attained by groin construction to retard loss of material from beaches, or to build beaches by trapping material. This latter is somewhat of a hen or egg complex, it not being always clear whether we build a beach by reducing the amount of material usually lost or by trapping material that usually would pass the area.

Unfortunately, we have not been able to formulate much that is done into a well supported regimen of reasoning. Many things are done successfully without our knowing why, which when tried in another environment fail miserably. It is to improve our understanding of these successes and failures, and to develop a sound basis for various commonly employed stratagems that we wish for more knowledge of ocean waves. It is to extend our possibilities for control of wave energy that we look to ocean wave research for both basic and applied research designed to eliminate the unknowns of the natural phenomenon.

I think it can be said that almost anything we discover about ocean waves will be of value to the student of shore processes. Probably the most pressing, immediate question is that of the transportation phenomena

associated with wave action. An urgent problem, with which I have been concerned, will illustrate what is needed. In many areas, notably the Great Lakes, a thriving industry supplying sand and gravel for construction from the lake bed has been built up. Many millions have been invested in the business. However, severe erosion of the lake shores has been linked in the minds of many with the dredging of material from the lake bed, with the result that the State of New York, for example, forbids Lake Ontario commercial sand and gravel dredging within its state limits. Ohio and Pennsylvania may be forced into some similar action. The question is this, does the removal of sand in relatively small quantities from the lake bed, in depths of 15 to 30 feet or more, result in recession of the shore line. We think the solution lies in a study of the transportation of material by wave action, how and where is it moved, what is the relation between wave energy and quantity moved, and finally, what is the ultimate effect on the shore line of the change in material balance caused by removal of the dredged material.

Another problem, which is of great economic importance, concerns navigable inlets or entrances on the coast. Most such inlets or entrances have been deepened or widened, or both, by dredging and require to be dredged periodically to maintain the desired widths and depths. Most of the material which shoals these channels moves, under wave action, along the coast to the entrance and into the channel. Jetties or similar structures may be built to obstruct the passage of sand into the channel. Ocean waves, in this case, not only move material into the channel, but also act on the structures, so their importance is two-fold.

Here again we need to know, quantitatively if possible, the relations between the material transported by wave action and the characteristics of the wave regime of the area, also the relations between the wave action and the structure characteristics required to insure stability and effectiveness of the structure. Until we know the characteristics of the changes in sea surface elevation and the associated internal movements of the water, we have little hope of solving our problem. The potential savings in more efficient design of structures are very large, a mere saving of 10 per cent in stone required on one given project having a money value of \$350,000.

It must be admitted that some applications of present knowledge of ocean waves are not widely recognized.

For example, in the fishing industry, many fishing days are lost by not taking advantage of the possibility of correlating operations with predicted sea surface conditions. Weather reports of high winds may result in cancelled operations because of supposed accompanying rough seas, when actual conditions would not preclude operations. More reliance on state of sea forecasts could well result in more efficient operations. Some acceptance of this procedure on the Pacific Coast has occasioned favorable reactions.

Forecasts of wave action in the Hawaiian area by the Weather Bureau

have been employed by the Corps of Engineers in breakwater construction activities with notable results. It has been possible to protect life and equipment from damaging wave action not otherwise suspected, as well as to maintain construction schedules in the face of apparently increasingly dangerous sea conditions, when forecasts of sea conditions indicated a safe upper limit for continuing work.

The possibilities of refraction and diffraction diagrams in furnishing information on wave characteristics and the effect of harbor structures are now being realized slowly. A more extended employment of such diagrams seems justified.

There are perhaps other applications of knowledge of waves derived from ocean wave research that escape me at the moment, but which would be apparent, were engineers more familiar with the results of ocean wave study. It is probable that the task of promoting familiarity must be assumed, at least for a time, by those conducting and reporting on ocean wave research. We must be scientists developing basic knowledge and engineers promoting the application of that knowledge to practical problems. If we are successful in such activity, we can expect strong support for our basic research and a large demand for applied research. Certainly, we must bend our every effort to insure the continuance of the progress we have made in the war years. The development and fostering of engineering applications of wave knowledge demand our attention not only because of their economic and technical importance, but as a means of insuring adequacy of support for basic research.

Discussion of the Paper

Dr. G. E. R. DEACON (*Admiralty Research Laboratory, Teddington, Middlesex, England*):

It is a very important part of the oceanographer's work to point out the applications of his discoveries to engineering and other sciences, but there comes a time when so many demands are made on him for the loan of apparatus and practical help of all kinds that he is in danger of having too little left for a concerted attack on the many fundamental problems he has to face. It becomes necessary to maintain a reserve of men and materials sufficient to allow such undoubtedly profitable diversions.

With reference to the use of wave knowledge by engineers, it may be regarded as regrettable that they have to spend so much of their time building enormous solid structures to reflect the wave energy which streams towards their country, and that there is no hope of building a breakwater which will absorb energy in a form which can be used.

DETAILS OF SHORE-BASED WAVE RECORDER AND OCEAN WAVE ANALYZER

By ARTHUR A. KLEBBA*

Woods Hole Oceanographic Institution, Woods Hole, Massachusetts

Research groups have been interested in measuring waves at some distance from shore. In order that this work might be furthered with minimum loss of time in securing data, we have developed a set of instruments capable of recording waves in depths of water to 150 feet. The actual recording takes place on shore and the pressure pick-up unit is connected by cable.

The installation is self-powered and can operate continuously or be set to record for fixed periods without attention.

The principal parts of the underwater unit are shown in FIGURE 1. This assembly is approximately 18 inches high and the pressure responsive element is the bellows A, in the upper part of the unit. As this is intended to perform under a static pressure of approximately 100 feet of water, it is held in extension by a comparatively stiff spring B. Bellows A forms a closed air path with another (bellows D), which is directly underneath it in the drawing. The second bellows is a relatively weak structure and will reproduce the motions of the first bellows. It can be seen that if a great difference in static pressures occurs, the lower bellows will be forced to the length of its travel. Compensation for changes in static depth and for tides is accomplished by installing a capillary tubing, C, to connect the air inside the bellows to the chamber immediately surrounding the bellows. This consists of a 6-foot length of copper tubing with a bore of approximately 17-thousandths of an inch. The actuating rod from bellows D is connected to move a coil of wire, E, at the lower end of the unit. This coil of wire is placed in a very strong magnetic field and small movements of the coil produce a voltage proportional to the speed of movement of the coil. These voltages would be the first derivative of the displacement taking place. However, it will be shown that the actual displacement is recorded. A two-conductor cable emerges from the unit and is connected to the recorder located on shore.

The principle of the electrical recording element is merely a description of a commercial fluxmeter galvanometer¹. This is a taut suspension galvanometer which has a small magnet attached to the lower end of the galvanometer coil. The orientation of this small magnet is to oppose the flux leakage caused by the main magnetic assembly. As a signal is applied to the galvanometer, it tends to rotate the coil. In the usual galvanometer, the torque of the suspension will cause an equilibrium position for a given

* Development of the instruments was supported by the Bureau of Ships, Navy Department. The author is indebted to Lawrence A. Thayer and Frank J. Mather for construction and held testing of the equipment.

voltage being applied to the galvanometer. In the case of this device, the small magnet is rotated so that a correction torque is applied in the direction of the rotation. This can be adjusted to be very nearly equal to the torque of the suspension and is adjustable by a field strength adjustment. The result of this compensation is to have, very nearly, a free drifting coil which will follow the motions of a search coil in a magnetic field. This results in a rotation of a mirror proportional to the movements of the exter-

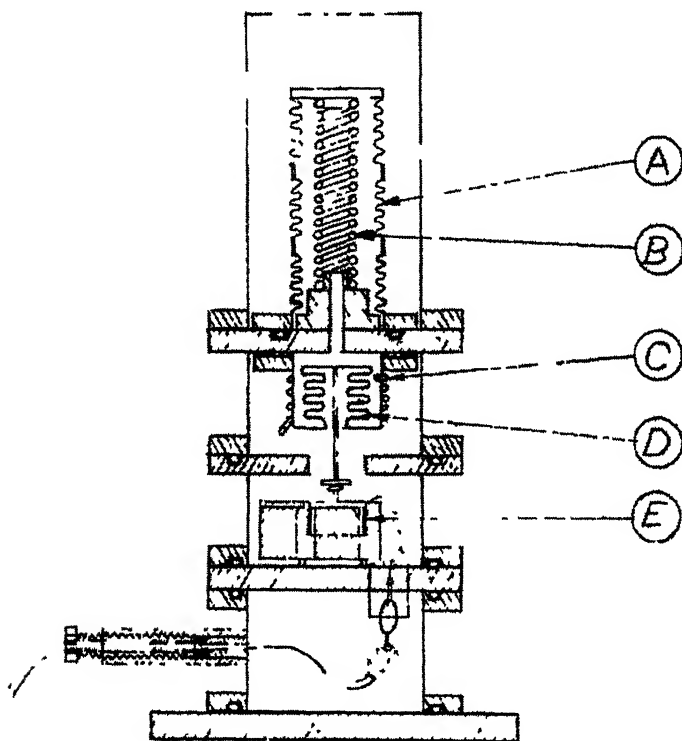


FIGURE 1

nal bellows of the underwater unit. In the commercial fluxmeter recorder,² a photoelectric follow-up system allows recording of ink on paper. The fluxmeter galvanometer has been incorporated in a recorder built at Woods Hole and is illustrated later. It was possible to install a Bourdon tube in the recorder so that simultaneous pressure changes and electrical indications of the pressure changes could be recorded. These are illustrated by FIGURE 2. The pattern at the top is produced by machine and illustrates a changing wave period from approximately $2\frac{1}{2}$ seconds up to 10 seconds.

The center and bottom patterns show how the electrical recording method follows the pressure changes. It can be seen that, in a few places, where essentially a static pressure has developed, the electrical recording tends

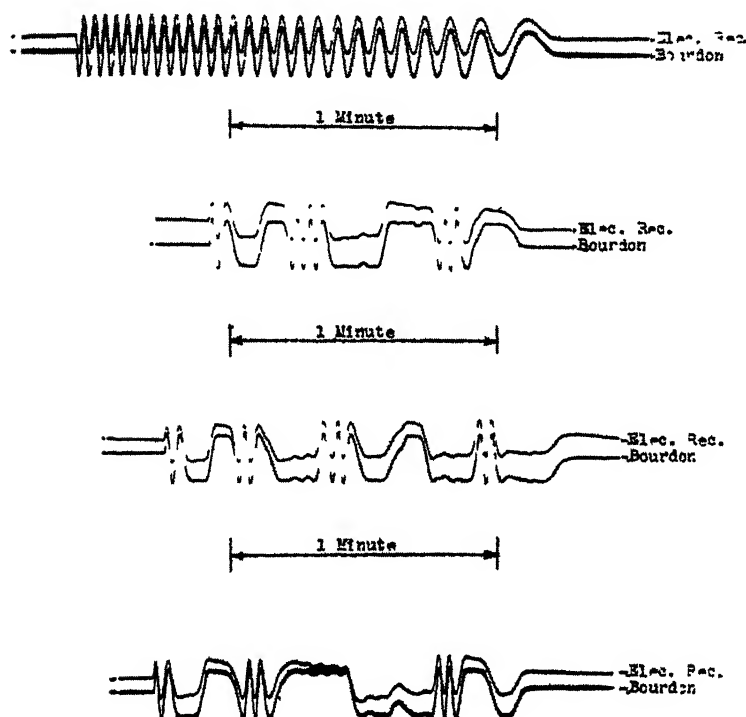


FIGURE 2.

to center. This is due to the fact that the fluxmeter galvanometer has the compensation device set to center slowly in the event that some disturbance has thrown it off scale. In addition, the slow leak has an effect here. The response to waves of various periods show it to be reduced to 90

per cent of its value, with a wave period of 44 seconds, compared to 100 per cent at 4 seconds.

The recorder, underwater unit, and power supply are illustrated in FIGURE 3. This is the recorder that produced the simultaneous records of FIGURE 2. The recorder weighs approximately 25 pounds and the under,



FIGURE 3.

water unit with ring weighs approximately 35 pounds. The power supply is an ordinary 6-volt storage battery. Approximately 1,000 feet of light rubber cable is illustrated. This equipment is convenient for measuring waves in protected areas and harbors. It can be set up in the space of an hour or two. An installation of this equipment is shown in FIGURE 4.

This is a large wooden box which has been fastened in place on the beach by stakes driven into the ground. In this particular case, the underwater unit was one mile off shore in approximately 85 feet of water. All the recording equipment is in this box and, after final adjustments are made, the box is closed and left to operate continuously or to take a 20 minute record every 3 or 4 hours. A set of records from this installation is shown in FIGURE 5. Each strip represents approximately 10 minutes of recording. This is a bottom pressure representation of surface waves and it can be seen that much manual work is required for extracting data.

It became of interest to find hidden components in ocean waves and we started the development of an ocean wave analyzer. British developments⁸



FIGURE 4

had preceded us in the field, thus the construction of an ocean wave analyzer was not original with us. As a first step in this direction, we developed a recorder to provide a record for mechanical analysis. The most simple record that would be useful in many ways was a half white, half black, record. The pressure pattern appears similar to a profile of the sea surface. The optical system of this recorder is shown in FIGURE 6. This is a plan view of the recording system. Lamp A is a line filament lamp with its filament in a horizontal position. Light from this lamp is collected by lens B and an image of the filament is produced at knife edge C. At this point a small portion of the light is cut off so that a cold filament end would not be reproduced on the photographic paper. This light is reflected by mirror D and is collected by lens E. After being reflected by fluxmeter mirror F and

mirror G, the image is focused on the recording paper. The portion of the image that corresponds to the cut off point at knife edge C is intended to fall near the center of the paper. This results in a bright line of light which has an abrupt halt in illumination. As the recording paper moves in an upward

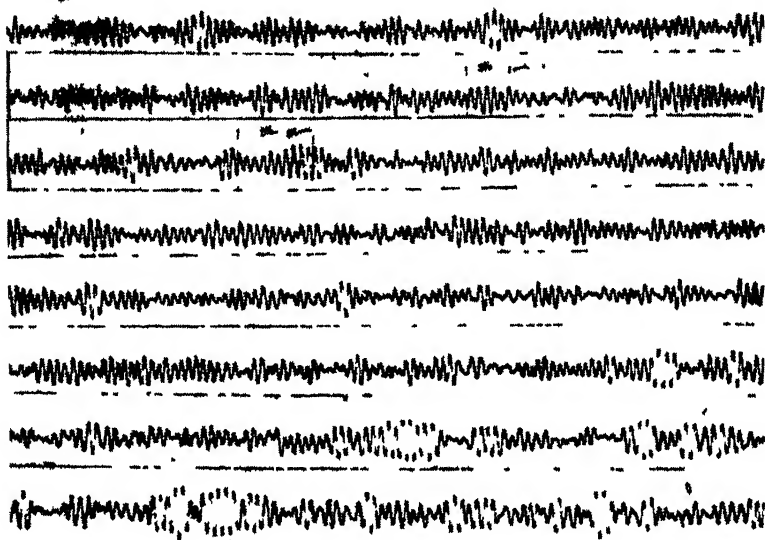


FIGURE 5

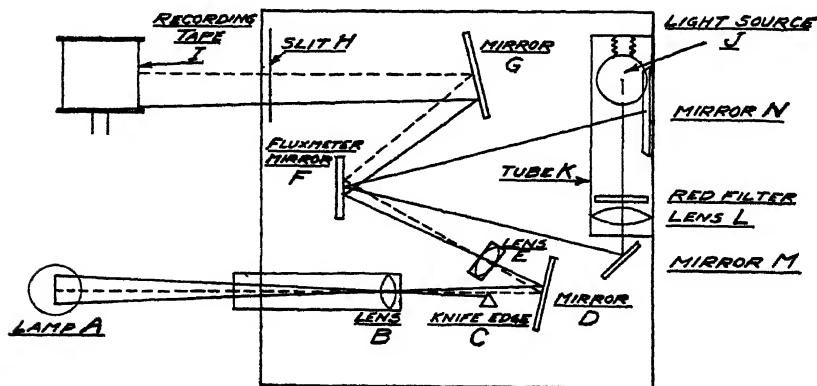


FIGURE 6.

direction, fluxmeter mirror F is rotated by the pressure fluctuations and the line of light moves across the face of the paper. As a result, half the photographic paper is exposed and a profile pattern is produced. Components J, K, L, M, and N are introduced to provide a visual filament image at a viewing slit, so that fluctuations can be observed.

We have added a control box, so that the program-switching clock and controls are contained in a unit separate from the recorder. The recorder is usually mounted by fastening a brass plate to the floor with long bushings and lag screws. The recorder has three legs with adjustments and this is placed over the brass plate. Locknuts are provided on each side of the brass plate so that positive and rigid leveling is obtained.

A few records from this instrument are shown in FIGURE 7. These were obtained off Cuttyhunk, Massachusetts. The width of the tape is 35 mm. and the timing marks are at 30 second intervals. The record is first received as a 200 foot roll of exposed tape. It is photographically processed and dried.

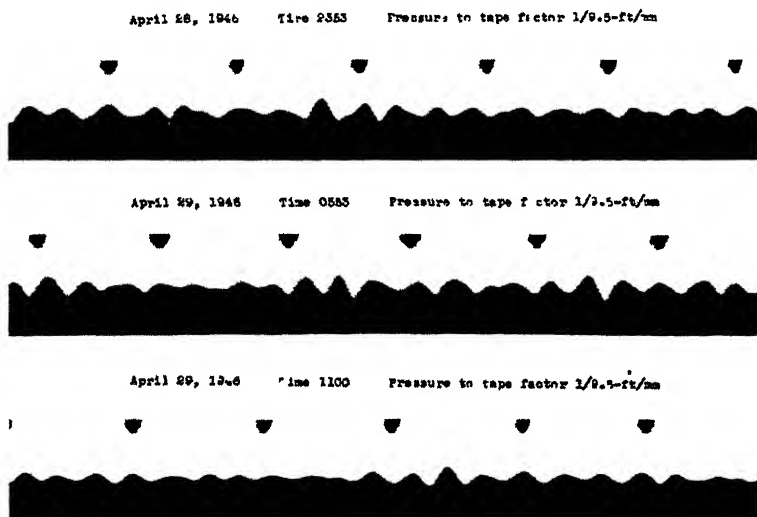


FIGURE 7

The analyzer for these tapes is shown in diagrammatical form in FIGURE 8. Operation is as follows: A 5 foot, 4 inch strip of record is placed around the circumference of a 20-inch drum, B. This is approximately 19 minutes of recording. Close to the circumference of the drum, shown by the block marked "photoelectric scanner," is a source of illumination, a lamp and a photoelectric cell. A portion of the record is illuminated and an image of the illuminated section is produced at a point in the box. At the point of the image, a set of slits is provided and unwanted portions of the record can be blocked out. The portion of the image desired passes through the open slide and falls on a photocell. As the drum is rotated, the photoelectric scanner sees the wave record as a bundle of frequencies. The signal from this portion of the circuit is sent through a fixed frequency filter. The frequency is approximately 125 cycles. The signal resulting from this filter is amplified

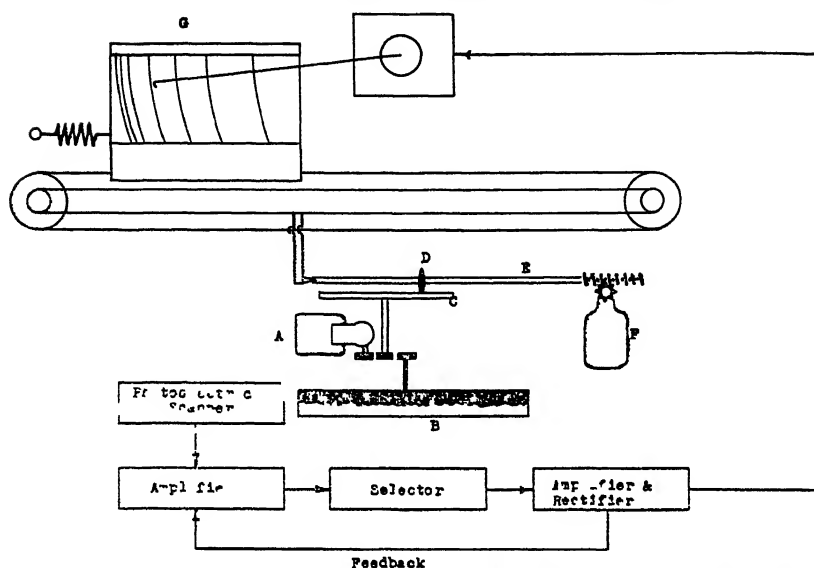


FIGURE 8 Ocean Wave Analyzer Mechanical Diagram

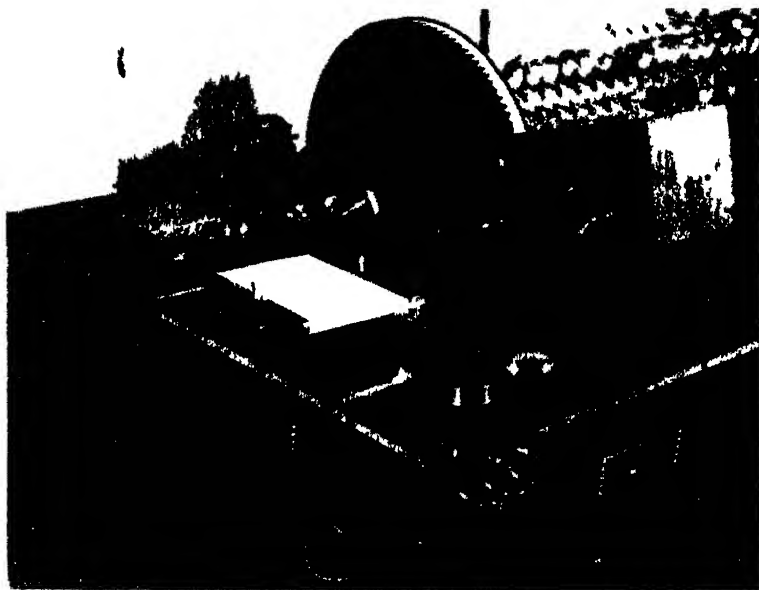


FIGURE 9

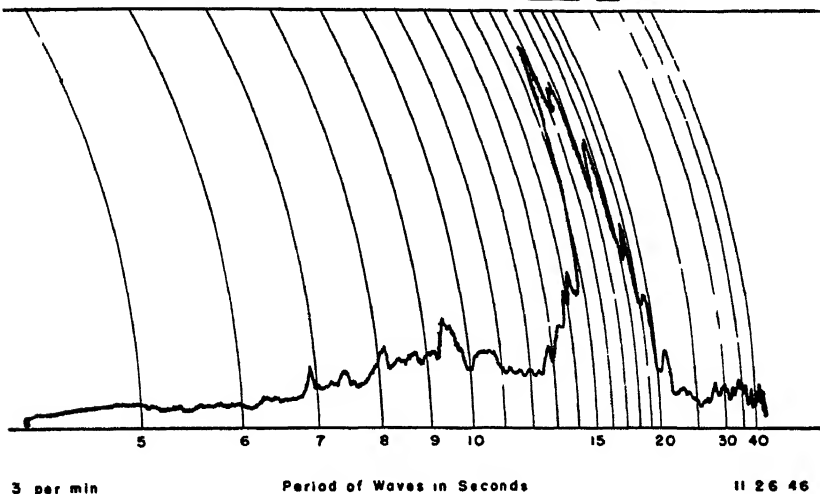
further and a portion of the signal is sent back to the last stage of the first amplifier so that better selection is obtained

The final output of the amplifier is rectified to operate the pen arm of the

WOODS HOLE OCEANOGRAPHIC INSTITUTION
WAVE FREQUENCY ANALYSIS

Date of record 9/15/46 Time 20 50 EST 5
 Depth 28' Record number 5374
 Pressure to tape factor ft/mm 1/2 F
 For reproduction of records only 1842534

Location Guffyhunk 41°24'12"N, 70°54'48"W
 Date of Analysis 11/4/47 by BH
 Max wave amp mm 45
 Surface wave 21 ft 14 Second

WOODS HOLE OCEANOGRAPHIC INSTITUTION
WAVE FREQUENCY ANALYSIS

Date of record 9/15/46 Time 20 50 EST 5
 Depth 28' Record number 538
 Pressure to tape factor ft/mm 1/2 F
 For reproduction of records only 1842534

Location Guffyhunk 41°24'12"N, 70°54'48"W
 Date of Analysis 11/4/47 by BH
 Max wave amp mm 6
 Surface wave 30 ft 1/25 Second

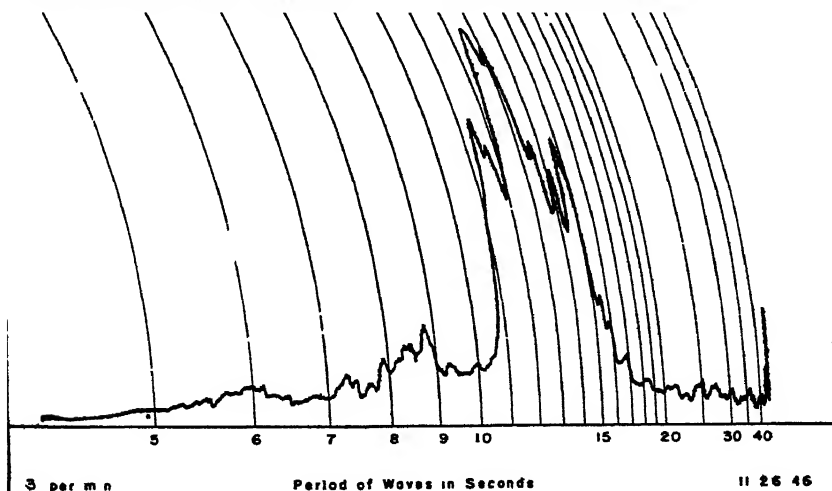


FIGURE 10

instrument. The rest of the device is to move the chart table to a position corresponding to various speeds of the drum. The wave spectrum is obtained by speeding up the drum and allowing it to decelerate slowly.

The chart table is moved in the following sequence of operations. Disc C, with motor A which is a brass disc, approximately 9 inches in diameter, rotates at the same speed as drum B. On the face of this disc, a small leather roller D rolls by friction contact. This is attached to shaft E, which is free to rotate, and the bearings are made so that it can move along its length. However, it is constrained in this lengthwise motion by a worm cut on the end of the shaft, which engages a worm gear attached to motor F. This is a synchronous motor. The action of the worm on the worm gear



FIGURE 12.

is such as to cause the shaft to progress to the left, as disc C rolls roller D. However, the action of the motor F is to move the shaft to the right. The constants are adjusted so that shaft E must rotate at 120 r.p.m. to remain at a fixed position. At low speeds, roller D is near to the circumference of disc C and at high speeds D is near the center of disc C. The end motion of shaft E is connected through a system of cables to the chart table B. The chart used is calibrated for particular recording tape speeds. A picture of the analyzer is shown in FIGURE 9. The drum B is shown in the background with a wave record around its circumference.

FIGURE 10 shows analyses of waves during a near hurricane, September, 1946, which passed off Nantucket, Massachusetts.

We intended to install some of our equipment at Bermuda and, in the fall of 1946, two of our field personnel went to St. Georges, Bermuda, and in months following, they installed an underwater unit southeast of Castle Harbor. The area concerned is shown in FIGURE 11. The underwater unit was first placed in approximately 130 feet of water and the cable was placed in Castle Harbor and Ferry Reach to allow recording at the Bermuda Biological Station. The length of the cable was approximately 5 miles. The underwater unit was identical to one shown in FIGURE 1. It was mounted to the top of a 450-pound concrete block with steel straps forming an apex over the unit. The block was approximately 2 feet square and 1 foot thick. Much difficulty was found in retrieving this anchor unit as it was found to be too heavy for this type of work. A later arrangement is shown in FIGURE 12. It consists of a frustrum of a cone, built of light steel rod covered with wire screen. In addition, several cable coils were placed so that the unit could be retrieved and a light cord fastening each coil of cable would be broken. In this way, the unit could be brought to the surface without actually raising any of the original cable.

References

1. RICH, T. A. 1940. *Memorandum on Fluxmeter Theory*. General Electric Company.
2. General Electric Company. Cat No. 32C120.
3. BARBER, N. F., F. URSELL, J. DARBYSHIRE, AND M. J. TUCKER. 1946. A frequency analyser used in the study of ocean waves. *Nature (London)* 158: 329-332.

UTILIZATION OF WAVE FORECASTING IN THE INVASIONS OF NORMANDY, BURMA, AND JAPAN

By CHARLES C. BATES*

Hydrographic Office, Navy Department, Washington, D. C.†

Introduction. War, grim destroyer of society, provided an unusual stimulus to ocean wave research in the years 1942 through 1946. Theory, observation, and prediction of sea, swell, and surf made the greatest strides in their history during this brief five-year period. The time already seems long ago when only the vaguest concept existed of the growth and decay of ocean waves and their ultimate transformation into breakers. Even in the early part of 1942, methods for observing and predicting these phenomena were essentially qualitative. The French had established a forecasting service for the open roadsteads of North Africa in the early 1920's, but results were still crude twenty years later. Naval meteorologists had done no better, and were accustomed to discussing wave conditions according to sea scales in which the terms might well mean different things to the man in a small boat and to the man aboard a battleship.

The groundwork for the quantitative forecasts of sea, swell, and surf required by large-scale amphibious operations was laid in both Great Britain and the United States during 1942. In that year, Instructor Commander C. T. Suthons, R.N., of the British Naval Meteorological Service, correlated a number of wave observations with wind conditions and prepared a memorandum which contained certain "rules of thumb" and crude forecasting graphs. About the same time, Mr. W. H. Munk in the Oceanographic Section, Directorate of Weather, Headquarters, Army Air Forces, was assigned the problem of developing wave generation and decay diagrams which could be used for the invasion of North Africa. For the same invasion, Commander R. Steere, U.S.N., prepared the first quantitative surf observation code. This code, named after its originator, was eventually revised and is now the currently used Combined Surf Code.

During early 1943, Mr. Munk worked with Dr. H. U. Sverdrup, Director of the Scripps Institution of Oceanography, to develop a technique of wave forecasting based on theoretical, as well as empirical, considerations. By July of that year, the necessary relationships had been established and a group of eight Army Air Force meteorologists, including the writer, was already assigned to study and test the embryonic forecasting method. Trial "hindcasts" made for the North African coast from the Northern Hemisphere Weather Map Series by this group indicated that the technique could be used by meteorologists after a relatively small amount of training. In

* Oceanographer, U. S. Hydrographic Office.

† The views expressed in this article are solely those of the author and are not to be construed in any sense as being the official viewpoint of the Navy Department.

fact, the short course in military oceanography built around the technique and taught at the Scripps Institution of Oceanography was cut in duration from the original three months to one month as refinements and organized material became available.

*Wave Forecasting for the Normandy Invasion**

When, in the fall of 1943, Lieutenant J. C. Crowell, A.U.S., member of the initially trained group, arrived in England and was assigned the task of preparing wave forecasts for the operations of the U. S. Assault Training Center at Woolacombe, Devon, many problems were still unsolved. These included such fundamental questions as the type of wave height the generating graphs provided, the manner in which waves were transformed into breakers while passing through shallow water, the effect on wave height when the waves approached the coast at an angle, the effect of opposing and following tidal currents, and the relationship between wave conditions and the efficiency of an amphibious operation. Answers to most of the problems had to be forthcoming within the next few months, for General Dwight Eisenhower, U.S.A., arrived in Great Britain during January, 1944, to organize Operation OVERLORD, the invasion of Normandy.

It was evident to all concerned that the success of Operation OVERLORD hinged upon the efficiency of the flow of troops and supplies across the beachheads. Even with favorable weather, it was estimated fifteen weeks or more would be required to transport as many Allied divisions across the English Channel as the Germans had available in Northern France and Belgium¹. To provide a satisfactory wave forecasting service for this operation, Captain L. G. Garbett, C.B.E., R.N., Director of the Naval Meteorological Service, collaborated with Colonel T. S. Moorman, Jr., U.S.A., Staff Weather Officer to the 9th Air Force and the 1st U. S. Army, in establishing the Swell Forecast Section, Admiralty. Organized with the approval of Commander R. Steere, U.S.N., aerologist to the Commander of U.S. Naval Task Force 122, the Section finally consisted of one British meteorologist (Instructor Lieutenant H. W. Cauthery, R.N.), two American meteorologists (Crowell and the writer), two American enlisted men (Technical Sergeant E. A. Lochner, A.U.S., and Sergeant E. L. Hynes, A.U.S.), and two WRNS ratings. Housed two floors underground, the Section had direct access to Admiralty's Forecast Section with its wealth of weather data and excellent communication facilities. The name, *Swell* Forecast Section, was not intended as a pun but rather as a security measure to direct thought away from the fact that the invasion might be scheduled for beaches not directly exposed to ocean swell.

* The writer wishes to point out that the results described in this paper were due to the combined efforts of the personnel of each of the wave forecasting groups with which he was associated. It is to be stressed that the paper should not be construed as a complete history of wave forecasting in the war theatres mentioned. There were several hundred meteorologists trained in the technique, and it is beyond the writer's scope to adequately and accurately describe their work in the field.

The objects of the Section were to develop the technique of forecasting sea, swell, and surf, and to provide forecasts on the basis of this technique for the invasion of Europe. The technical problems facing the Section were four in number; (1) to forecast the height and period of ocean swell coming from the Atlantic; (2) to determine the extent to which this swell would penetrate the English Channel; (3) to forecast the height and period of waves caused by local winds in the Channel; and (4) to study the effects of shallow water, tidal currents, and coastal irregularities on waves. The final aim, of course, was to forecast surf heights on specific beaches. The investigations were carried out under two handicaps, security and limited time. Although the date, place, and other details of the impending operation were known to the Section, security measures were so rigid that wave researchers in the States could not be contacted concerning the problems mentioned earlier. As only three months were available, the empirical approach was adopted as the one most likely to provide the necessary information.

The first task undertaken was the organization of a synoptic network which eventually totaled fifty-one wave reporting stations (FIGURE 1), and was probably the largest of its kind ever organized. In the main, the stations were His Majesty's Coast-guard lookouts, manned by retired seamen. Visual observations made at these stations were reported in a seven-figure group, IHHPPM, where II is the station number, HH the average wave height in feet, PP the average wave period in seconds, and M the difference in feet between the height of the maximum wave observed and the average height. Daily observations were made at 0700, 1300, and 1800 GMT for an interval lasting three minutes and included a wave count and an estimate of the height of each wave breaking during that time. Rocks or other objects of known size occurred in the surf zone at a few of the stations and aided in the estimation of height. As soon as observations were made, the data were phoned to district headquarters and then relayed as a collective by teleprinter direct to the Admiralty. To study the exposure of each location and to provide the necessary instruction, all the stations were visited by one or more members of the Section. Although visual observation is subject to considerable error, the close spacing of stations permitted checks of one against the other, and the large amount of synoptic information supplied was invaluable for verification and research purposes.

Four of the stations (Padstow, Pendeen, Weymouth, and Newhaven) had aneroid pressure-recorders installed in comparatively deep water by the Director of Mine Design. Reports from these stations, made in the same fashion as those from the visual stations, were extremely valuable in checking the visual reports and in providing data on waves in deep water.

A special reporting station was also established at Weston Mouth. This location was chosen because it is protected from Atlantic swell and the

maximum fetch available is similar to that of Seine Bay, the site of the assault beaches. The station consisted of seven dan buoys, topped by graduated poles, and laid at equal intervals along a straight line to seaward. Observations were made by sailors five times daily with the aid of a graticuled theodolite and stop watch. However, the unusually calm spring weather prevented the achievement of the purpose of the station, *i.e.*, the acquisition of a large group of reliable observations on the transformation of

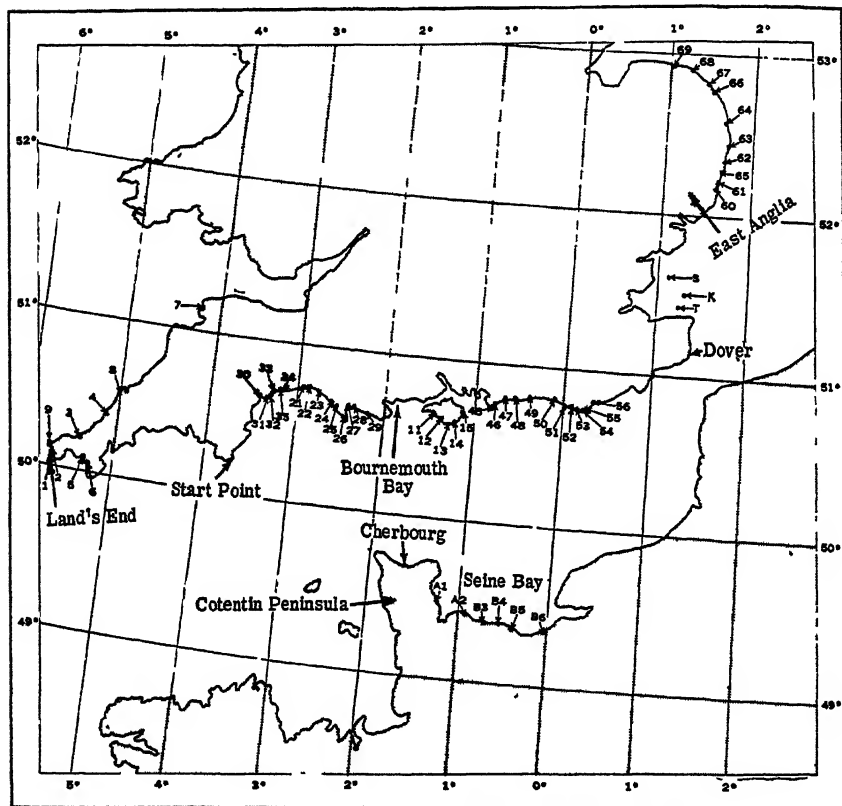


FIGURE 1. Index map showing location of English wave reporting stations and the invasion beaches of Normandy (for names of stations and beaches, see TABLE 1).

waves moving into shallow water. In fact, the station reported waves one and one half feet or higher on only forty-two occasions in four months of operation.

Wave information of lesser value was received from three anti-aircraft "forts" in the Thames estuary, from high-speed motor launches on night patrol out of Dover, and single-engined aircraft (P-51s) flying weather reconnaissance over the Channel. Checking of wave conditions on the

assault beaches proper was carried out by studying aerial photographs of the beaches. This was accomplished by determining the depth of breaking from these photographs by referring to special tidal data and beach profiles prepared for the invasion and then converting depth of breaking into breaker height by assuming breaking occurred in water one and one half times deeper than the height. This method, although crude, indicated no greater error in the forecasting technique being employed.

TABLE 1
WAVE REPORTING STATIONS AND ASSAULT BEACHES

Num- ber	Station	Num- ber	Station
1	Sennen	45	Selsey Bill
2	Cape Cornwall	46	East Beach
3	Go lreivy	47	Bognor Regis
4	Fistral Beach	48	Littlehampton
5	Prah Sands	49	Worthing
6	Loe Bar	50	Newhaven
7	Woolacombe	51	Kemp Town
8	Constantine Bay	52	Cuckmere Haven
9	Pendeen	53	Birling Gap
11	Blackwood Point	54	Holywell
12	Atherfield Point	55	Eastbourne
13	St. Catherine's Point	56	Pevensy Bay
14	Woody Point	60	Sizewell
15	Foreland	61	Dunwich
21	Seven Rock Point	62	Benacre
22	Charmouth	63	Lowestoft
23	West Bay	64	Gorleston
24	Abbotsbury	65	Southwold
25	Langton Herring	66	Polling
26	Fortuneswell	67	Mundesley
27	Osmington	68	Cromer
28	Osmington Mills	69	Cley
29	Lulworth Cove		<i>Assault Beaches</i>
30	Orcombe Point	a-1	Utah, American
31	Budleigh Salterton	a-2	Omaha, American
32	Ladram Bay	b-3	Gold, British
33	Sidmouth	b-4	Juno, British
34	Beer Head	b-5	Sword, British
35	Weston	b-6	Band, British

Fort Reports: S, Sunk; K, Knock; T, Tongue.

Early forecasting of the Section was directed toward predicting swell conditions at Land's End. A fair degree of success was achieved, although the rigid accuracy desired was not obtained in all cases. Work was also initiated into determining how far swell penetrated the upper reaches of the English Channel. Cornish² and various coast guard personnel had mentioned that swell occasionally penetrated at least as far as Bournemouth Bay. After studying the observations of the wave reporting stations, it was found that long-period waves rarely appeared at stations east of Start Point. By analogy, it was to be expected that swell approaching the

assault beaches from south of west would be cut off by the Cotentin Peninsula and that swell from a direction close to due west, even though of appreciable height off Cherbourg, would be considerably reduced in height at the beaches. From this and certain theoretical considerations, it was decided that only extremely high swell from the west would need to be taken into account in forecasting surf heights on the beaches in question and that such waves would diminish to less than half their original height. This conclusion was verified on October 14, 1944, when the first really noticeable swell appeared at Omaha Beach with breaker heights of 2 to 3 feet and a period of 9 seconds, although heights of 12 to 15 feet were reported off Cherbourg. During the summer months, such a situation was unlikely and the problem resolved itself into forecasting waves and surf resulting from local winds.

Fetches in the central part of the Channel and at the assault beaches were rarely greater than 120 miles. Verification of forecasts soon indicated the tremendous importance of correct wind forecasts. To provide accurate wave forecasts, *i.e.*, those with an error of one foot or less in height if the heights were less than five feet, and with an error of 2 feet or less if waves were greater than five feet in height, wind forecasts had to be in error less than one Beaufort force and less than two points ($22\frac{1}{2}$ degrees) in direction. Forecasts of wave period were unnecessary because variation in the slope and period of wind waves did not appear to affect small craft operation noticeably.

Before surf heights could be predicted, it was necessary to estimate the effect of shallow water and coastal irregularities upon deep-water wave height. As mentioned before, a comprehensive theory of these effects was not available, and it was decided to obtain a direct relationship between observed breaker heights and wave heights computed from generation diagrams. Such a technique would also incorporate correction factors needed to take into account tidal stream effects, influence of land bounding the channel, error in determining wind speed from meteorological charts, possible errors in the generation graphs, and several lesser considerations. If the necessary relationships between computed and observed data could be established for winds blowing onto the English coast, it was hoped that similar relationships would exist for winds blowing onto the French coast.

It was also necessary to determine how the height values extracted from generation graphs might be compared to observations which reported both average and maximum wave heights, particularly since the maximum height reported was often double that of the average height. It appeared that small craft operation was concerned neither with the occasional maximum wave nor with the average value, which is considerably depressed by the many small waves present in a wave train. However, a value half way between the average and maximum height appeared to be highly useful. This value, termed the "Predicted Height," was used for purposes of com-

parison. 'The value falls amazingly close to the "Significant Wave Height" defined by Sverdrup and Munk about a year later.'³ (See FIGURE 2).

Although a complicated shoal system existed off part of East Anglia, the ten wave observation stations spaced four to fourteen miles apart along that coast proved particularly valuable in working out the relationship between computed and observed height. FIGURE 3 illustrates the results of a study the writer made of heights observed during nineteen synoptic weather situations in which the wind direction held with 15 degrees of a given direction for eighteen hours or more during the months April-July, 1944. The hatched areas of the illustration are the zones in which weighted values for over two hundred height values occurred, using both Sverdrup-

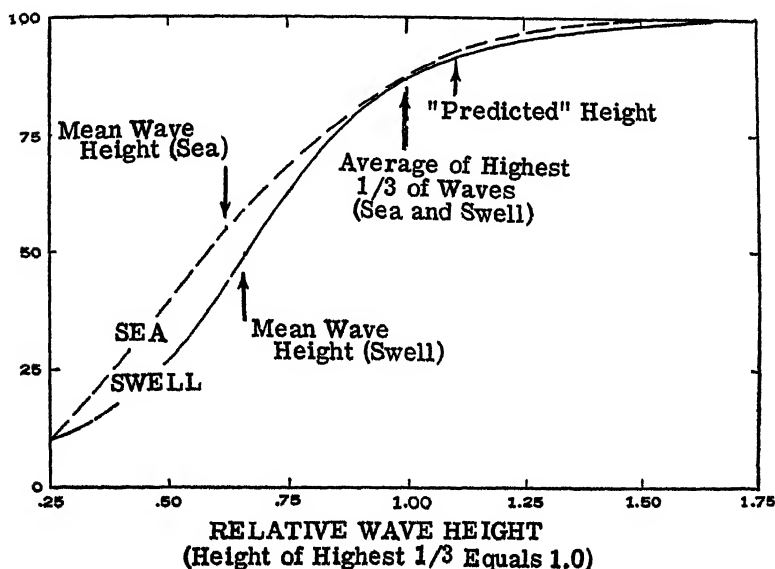


FIGURE 2. Per cent of occurrence of relative wave height for sea and swell (after Scripps Institution of Oceanography's Wave Report No. 68).

Munk⁴ and Suthons generation graphs. It is evident that, for waves approaching along the normal to the beach, the Sverdrup-Munk values were a trifle high. This discovery caused the Swell Forecast Section to apply a reduction of 10 per cent to values computed by this method to obtain breaker height. The computed Suthons values were much higher than the observed heights. Because of this, and because the Sverdrup-Munk curves were presented in a much more usable fashion, the latter curves were used in all operational forecasting. Other theoretical and empirical considerations likewise indicated that the increase in height of wind waves at the time of breaking could be neglected, a fact substantiated by later wave research.

Because the East Anglian stations had an accumulative exposure of about

268 degrees, the observations provided a clue as to the extent heights were reduced when waves approached the shore at an appreciable angle. The spread of values for different angles of refraction is shown in FIGURE 3,

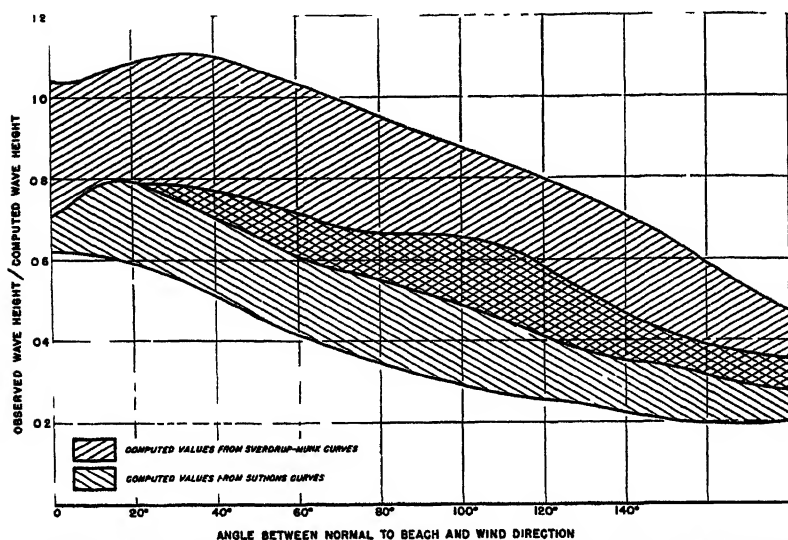


FIGURE 3. Variation of wave height with angle of approach to shore, according to observations made along the East Anglian coast.

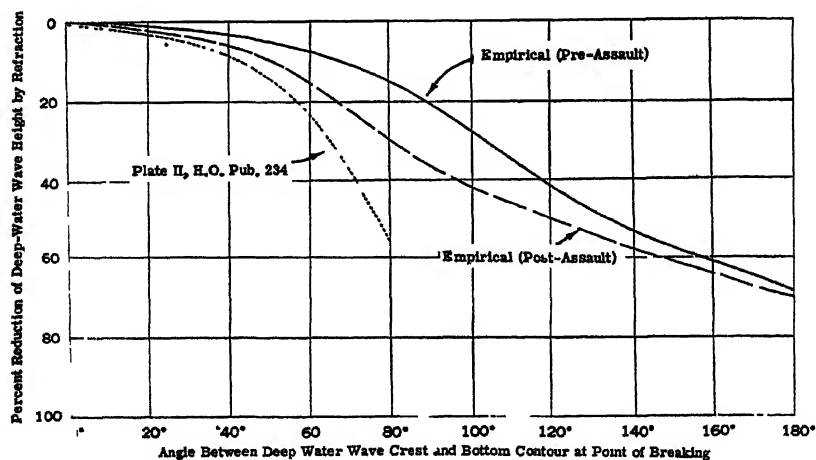


FIGURE 4. Values for reduction of wave height by refraction according to theory and two sets of empirical data.

while the values used in the Swell Forecast Section are given in FIGURE 4. It will be seen that the values used are in good agreement with the theoretical values given for a simple beach in the subsequently published "Breakers and

Surf, Principles in Forecasting⁷⁵ for refraction angles less than forty degrees, but that considerable discrepancy occurs for greater angles.

The empirical values obtained for angles of refraction between ninety and one hundred and eighty degrees were particularly valuable because the technique of preparing wave refraction diagrams as developed by O'Brien and associates⁶ was unknown to the Section at that time.

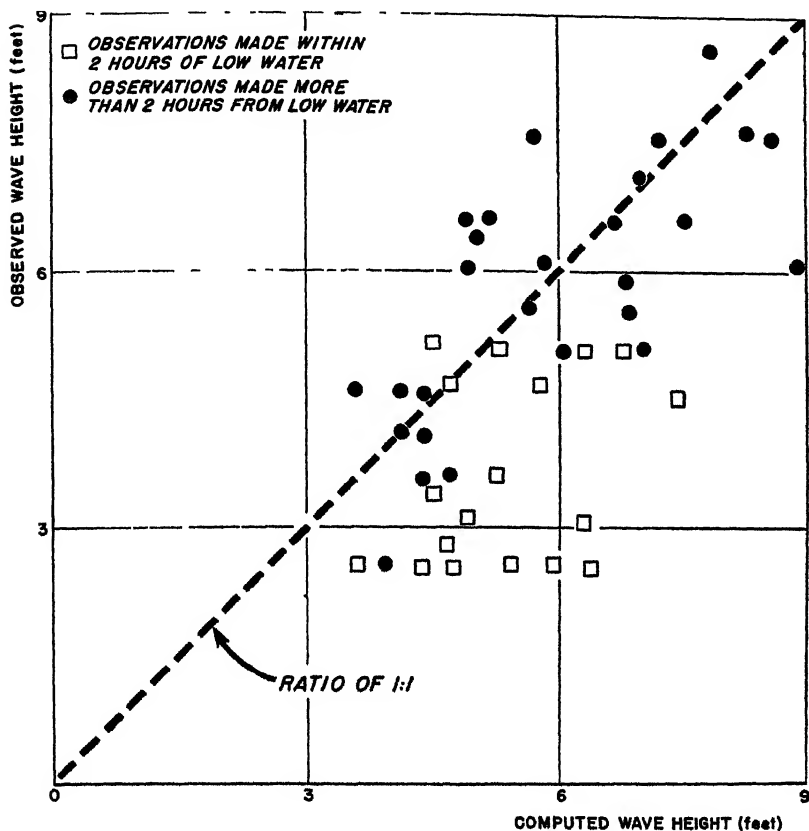


FIGURE 5 Computed versus observed wave heights for different tidal stages at Cley, Norfolk.

A third interesting feature found in the East Anglian data was that heights observed at certain stations within two hours of low water were inexplicably lower than the computed heights. This reduction, which averaged about twenty per cent, was particularly well evidenced in the distribution of height values for Cley, Norfolk, where the observations made more than two hours from low tide center about the ratio of 1:1 for observed versus computed heights, but the observations within two hours of low tide have only one value with a ratio greater than 1:1. (See FIGURE 5). Investigation

indicated that there should be no appreciable error in observing at low tide, because the water line retreated only one hundred and fifty feet or so in most cases. The irregular offshore bottom topography likewise appeared to be incapable of causing the unusual decrease. In fact, one of the stations with a relatively smooth offshore profile had a greater height reduction than stations with irregular profiles. A third possibility was that strong tidal streams opposed the forward motion of the waves and caused a pronounced loss of wave energy offshore through partial breaking. This effect is believed to be the cause, inasmuch as this is a region in which the time of low water coincides roughly with the time of the strongest northerly-flowing tidal streams. Since these tidal streams range from 1.0 to 1.5 knots at neaps and from 2.0 to 2.5 knots at springs, and most of the situations studied were for waves approaching from the north, it is highly probable that the tidal streams acted as a partial breakwater. This phenomenon of opposing tidal streams acting like a breakwater has long been mentioned in the literature, one of the more remarkable examples being reported in the Shetlands by Stevenson⁷ during the last century.

The tidal current effect on wave height was given additional scrutiny because one of the basic publications on English Channel weather maintained that the sea conditions raised by winds in the Channel depended largely upon the tidal streams. However, the wave observations along the Channel coast did not show any undue fluctuations in height, as were observed in East Anglia and as one might be led to expect from this statement. Although several stations did report suspicious heights, investigation indicated that the water line receded a considerable distance at these stations and reliability of observations made during low tide was open to considerable question. It was decided that tidal streams never raised the breaker heights if tidal streams opposed wave motion offshore. Tidal streams off the assault beaches generally paralleled the shore such that the effect could be ignored in forecasting surf conditions there. However, it was recognized that strong tidal streams, such as those found off the northeast tip of the Cotentin Peninsula, would cause marked changes in deep-water wave characteristics. Before D-Day, the Section believed that only the wave steepness would be markedly changed, the height increase being of the order of a foot or less. Subsequent work at Scripps proved this assumption to be in error as it was found that the height increase could be two or three times the original deep water height if the opposing current were strong enough and breaking did not occur.

When in early April, 1944, the Supreme Headquarters, Allied Expeditionary Forces (SHAEF), required 5-day wave forecasts for the English Channel and adjoining sea areas, the Swell Forecast Section had acquired a considerable backlog of experience in handling the problems outlined here. Preparation of the OVERLORD weather forecasts was the result of frequent conferences held *via* private secure telephone between three weather

centrals (Air Ministry, U. S. Strategic Air Forces, and Admiralty), and the three staff meteorologists representing SHAEF, Allied Air Headquarters, and the Allied Naval Commander, Expeditionary Forces. After the general synoptic picture for the next five days had been decided upon in each of the weather conferences, the conferees agreed on the wind forecast to which one of the wave forecasters had been listening at Admiralty. Then, either alone or in consultation with another wave forecaster, the wind-wave forecast was prepared and incorporated with the swell forecast made just prior to the conference, in order that the complete wave forecast could be given at the close of the conference.

From June 15 onward, forecasts for SHAEF were made daily for three-day periods, while Allied Naval Commander, Expeditionary Forces, received forecasts twice daily for a two-day period. Forecasts were also sent *via* radio to USAAF mobile weather stations in Normandy to aid in preparation of forecasts used in unloading operations, and to the Office of the Chief Engineer, European Theatre of Operations, for inclusion in the Daily Hydrographic Bulletin issued by that activity to the U. S. Ground Force Commands.

The rapidity with which the wave forecasts had to be prepared and the control exercised on fetch and refraction values by the complex coastal outline caused the Swell Forecast Section to develop "Surf Prediction Diagrams" for beaches with different characteristics, namely the British group, Omaha, and Utah. The diagram for Omaha Beach is shown in FIGURE 6. The polar section of the diagram delineates the maximum effective duration for various wind speeds from any direction, with the dashed parts of the curves indicating indirect fetches where waves were likely to appear because of refraction and diffraction processes even though headlands intervened. A comparison was made between the maximum effective duration determined from the diagram and the actual duration. The lesser value was then entered along with the proper wind speed in the generation graph based on the Sverdrup-Munk curves. The wave height so determined was reduced by the appropriate correction factor given in the rim of the diagram, with the resulting value being considered the predicted surf height.

The accuracy of operational forecasts made in the above manner is of interest. Using the definitions given earlier for accuracy of wind and wave forecasts and excluding all cases in which wind forecasts were in error, it was found that wave forecasts were correct eighty-eight per cent of the time for the assault beaches in the period June 6-30. If all bases are excluded in which winds were light or directly offshore, because no forecasting skill is involved, the forecasts were correct eighty-three per cent of the time for the same period.

In addition to information provided by the Swell Forecast Section, "on-the-spot" climatological studies and forecasts of wave conditions were

supplied field commanders during the pre-assault and assault phases by staff meteorologists such as Steere with Task Force 122 and Lieutenant D. W. Pritchard, AUS, with Headquarters, 1st U. S. Army. These meteorologists had been supplied with special "Notes on the Sea, Swell, and Surf in the English Channel" and pertinent surf prediction diagrams prepared just before D-Day which incorporated most of the "know-how" acquired during the four months of the Swell Forecast Section's existence.

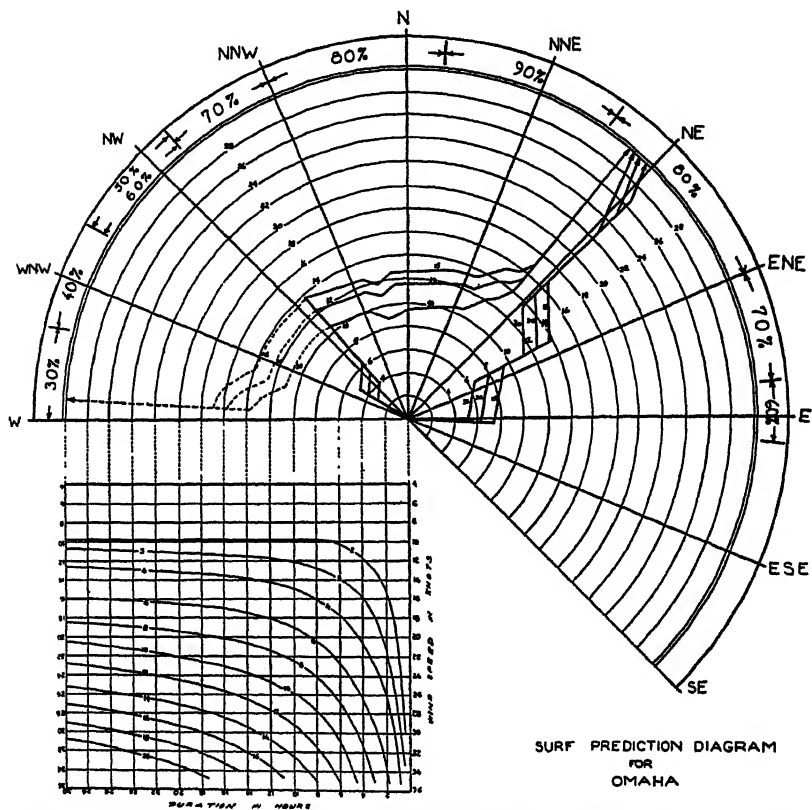


FIGURE 6. Surf prediction diagram as prepared for Omaha Beach a month before the Normandy invasion.

Although the invasion was originally scheduled for June 5th, the passage of a cold front caused a 24-hour postponement of D-Day. During the early part of the 6th (D-Day), wind waves offshore still averaged three to four feet in height with occasional heights from interference as high as six feet. All the beaches with the exception of Utah were directly exposed to winds of 12 to 18 knots, which raised three to four foot surf throughout the day. On sheltered Utah, surf heights were two feet and less. This condition remained until the afternoon of June 7th, when the northwest winds dropped to 5 to

10 knots and permitted waves to be but 1 to 2 feet high on all beaches by evening.

For the next eleven days, wave conditions did not markedly hinder invasion operations. On June 19th, however, a steep pressure gradient between a large anticyclone, centered northwest of the British Isles, and a comparatively weak cyclone, spreading across France and Spain, caused strong northeasterly winds in the Dover Strait and the English Channel. Although the naval command ships, USS AUGUSTA and HMS SCYLLA, only

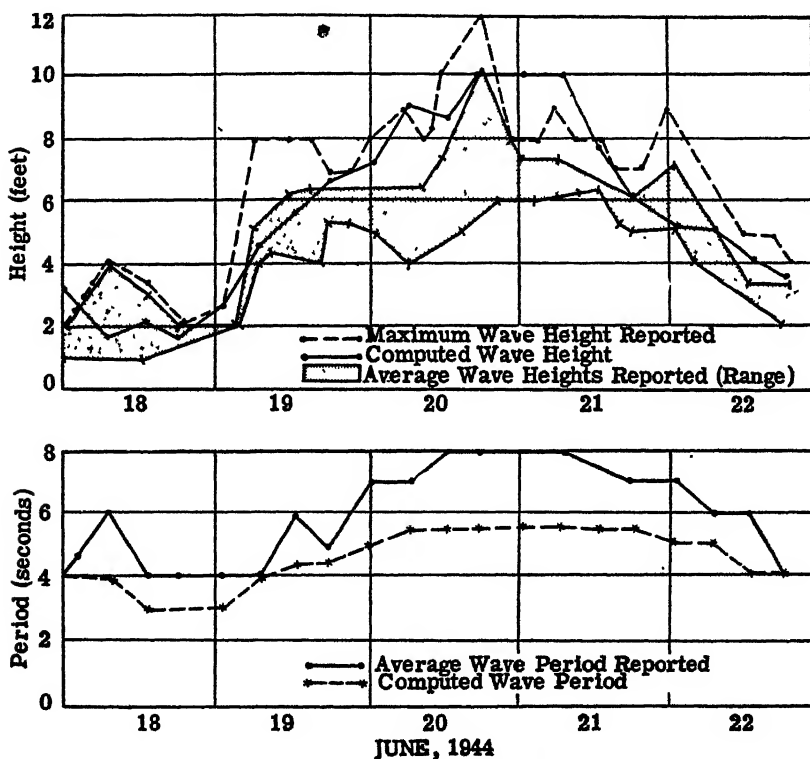


FIGURE 7. Waves observed and computed ("hindcast") for the area immediately off the Normandy Beaches ($49^{\circ} 26'N$ - $49^{\circ} 30'N$; $00^{\circ} 30'W$ - $00^{\circ} 50'W$).

reported average wind speeds of about twenty knots and maximum wave heights of ten to twelve feet off the beaches, this episode was called the "Big Storm," for seven hundred landing craft were damaged or lost and the floating Loebnitz piers on Omaha Beach wrecked beyond repair⁸ (See FIGURE 7). The twenty-four hour forecast for the 19th called for an increase in wind strength sufficient to raise wave heights to four feet, but heights of six to eight feet actually existed by that evening. A revision prepared a few hours after the routine forecast, however, provided sufficient warning, if a

storm plan had been placed in effect. Forecasting for the remainder of the storm was particularly accurate, the drop in wave heights to operational values being forecast within two hours of the actual time of occurrence.

It is interesting to speculate on what might have occurred if D-Day had been postponed to the next favorable tidal period which happened to fall just before the "Big Storm". With the large high-pressure cell centered just west of Great Britain, the temptation was to forecast light winds instead of the packing of the isobars and resulting strong northeasterly winds which actually occurred over the Channel. If such had occurred, the story of the buildup phase on the beachhead, by coming just at the time of the "Big Storm," would have been much different than it actually was. This particular meteorological type was troublesome throughout the summer, for it re-appeared several times, though in a less pronounced fashion. Even if the northeasterly winds were light during night and morning hours and wave heights correspondingly low, reinforcement by the afternoon sea-breeze quickly raised heights to four feet or above, a value sufficient to shutdown DUKWs transporting cargo to shore from the freighters anchored outside the artificial breakwaters. As winds reported by weather stations just a few miles inland were generally light and variable on such occasions, forecasting from the beachhead itself was the only practical solution to the problem.

If, because of lack of information on surf conditions, work was suspended even one hour earlier than necessary, it resulted in the loss of hundreds of tons being off-loaded. In a similar fashion, any delay in resumption of operations after a siege of high waves cost valuable operational hours. To reduce this loss, British and American beachhead commands utilized both the forecasts from Admiralty and special forecasts prepared locally, the latter being more up-to-date because of slow communications between England and Normandy. The Flag Officer, British Assault Forces, had his own mobile weather station under Lieutenant J. H. C. Fulford, R.N.V.R.; Chief Warrant Officer White, U.S.N., at Cherbourg provided U. S. naval beach activities with necessary information; and Lieutenants D. W. Pritchard, AUS, and R. O. Reid, AUS, issued wave predictions required by the U. S. Army Commands. After the breakout, it was evident that the American beaches required a weather station of their own and could no longer rely on weather units at air fields. Under the command of Pritchard, Detachment YK of the 21st Weather Squadron, 9th Air Force, issued its first forecast on this new basis during September 12, 1944. As operations over the beaches did not cease until mid-November, the beach commands became increasingly weather conscious. Work initiated by the writer, after a visit to the beachhead in July, indicated a relationship so pronounced between wave height and the amount of tonnage discharged that wave forecasts actually became quantitative estimates of the scale of the next day's activities (FIGURES 8 and 9). This work was then developed by

Pritchard and Reid at Detachment YK on Omaha Beach. Seiwel's recent paper in the *Military Engineer* discusses in detail the work of this detachment,⁹ therefore, further discussion of this phase is omitted here.

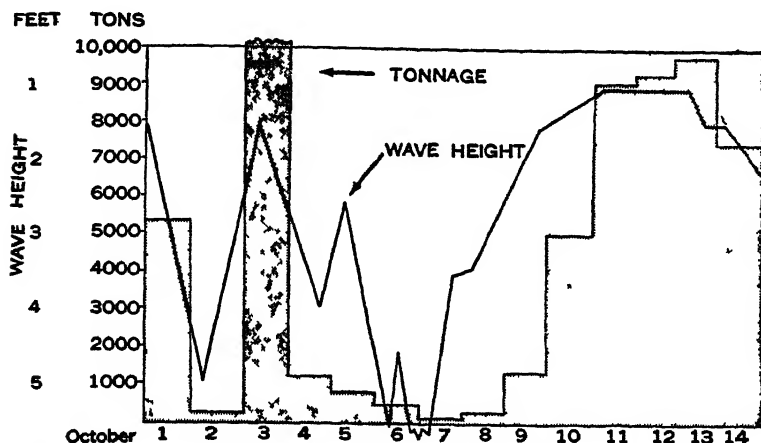


FIGURE 8. Effect of wave height upon amount of tonnage unloaded daily at Omaha Beach, Normandy during the first two weeks of October, 1944

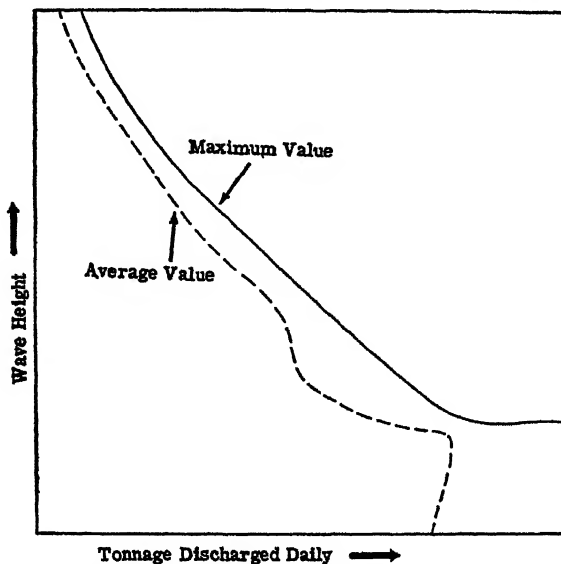


FIGURE 9. Relation between rate of unloading and the wave height at Omaha Beach, Normandy

Forecasting for Amphibious Operations in the Bay of Bengal

Faced with the necessity of taking landing craft across the Bay of Bengal during the southwest monsoon, the Royal Navy transferred the Swell

Forecast Section, Admiralty, to the Joint Meteorological Centre, Colombo, Ceylon, in late 1944, to supply wave forecasts for the Southeast Asia Command and the British East Indies Fleet. Through the courtesy of the American Army's weather service, the key American personnel (Crowell, Lochner, and the writer) remained with the unit.

The problems facing the Swell Section, as it was called following the transfer, were as fully numerous and intriguing as those which had faced the group before OVERLORD. Tropical meteorology did not possess methods of determining wind speed and direction with the accuracy desired in wave forecasting. The strong southwesterly winds of the monsoon were predominantly offshore for the area in which synoptic weather reports were available, but onshore in the assault areas, where there was little synoptic weather information. Fetches were now of the order of hundreds and thousands of miles, rather than the tens of miles found in the English Channel. The entire Bay of Bengal was exposed to southerly swell that could originate anywhere in a vaguely observed oceanic expanse stretching southwards as far as Antarctica. Forecasting the amount of height increase resulting from interference between different swell trains was also considered an outstanding problem after it was observed that breakers averaging three to four feet in height at Colombo were sometimes accompanied by maximum heights of fourteen feet.

As in England, the first work accomplished was the organization of a wave reporting network. This time the synoptic net composed of twenty stations was over three thousand miles in length (FIGURE 10 and TABLE 2). Two reports were made daily of average and maximum breaker height, average period, and the angle which the breakers made with the beach. During the organization of this network in February and March, 1945, each of the stations, with the exception of Cocos and Diego Garcia Islands, was visited by a member of the Swell Section for instructional purposes. Because of the nature of travel in Asia, these visits were often extremely interesting. For example, when Crowell visited Saugor Island, in the delta of the Ganges, he found himself the center of attraction for the entire populace, since he was the first white man to enter those parts in many years. The Admiralty Research Laboratory, Teddington, England, also cooperated in the observation program by sending two technicians, Mr. Alexander and Lt. Ogilvie, RNVR, to install and maintain two pressure-type wave recorders. In order to determine whether there was a suitable installation site along the Arakan coast, Crowell flew behind the Japanese lines in a light airplane. He decided that shallow offshore profiles and living problems eliminated this possibility, even though British troops were advancing rapidly enough to secure any site selected. Plans were then made for installing the instruments at Colombo and Addu Atoll, and this was accomplished during the summer of 1945.

Although the Swell Section knew that the initial landing for which fore-

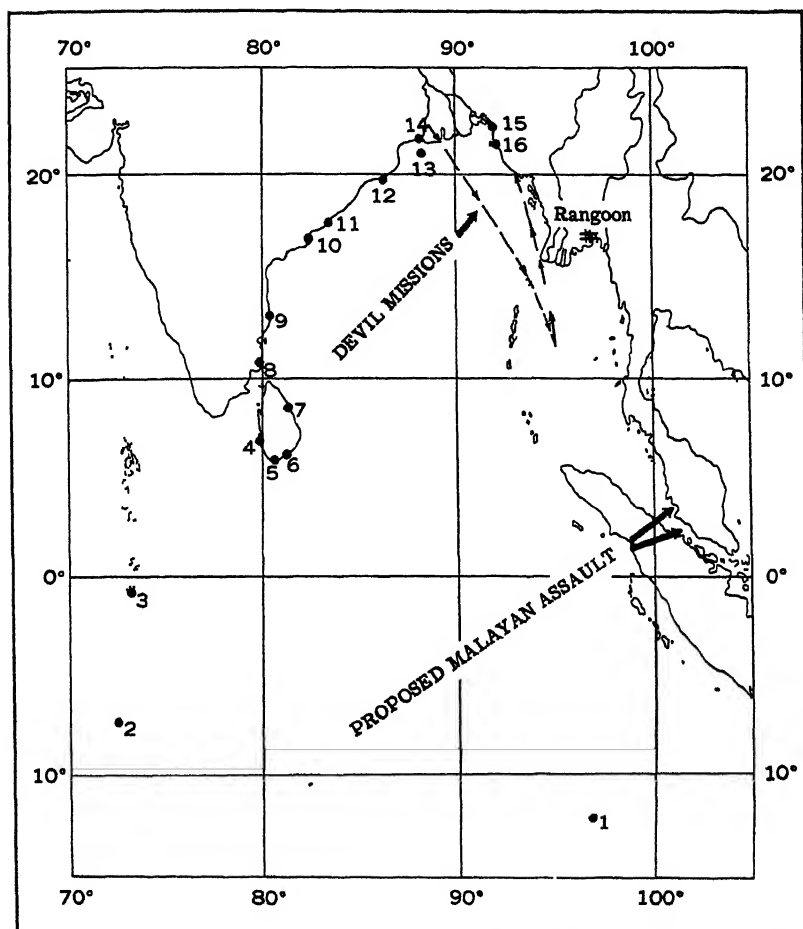


FIGURE 10. Index map showing location of wave reporting stations established to aid wave forecasting for the Bay of Bengal (for names of stations, see TABLE 2).

TABLE 2

WAVE REPORTING STATIONS ESTABLISHED TO AID WAVE FORECASTING FOR BAY OF BENGAL

Num- ber	Station	Num- ber	Station
1	Cocos Island	9	Madras (two)
2	Diego Garcia	10	Cocanada
3	Addu Atoll (two)	11	Vizagapatam
4	Colombo	12	Puri
5	Kogalla	13	Sandheads Lightship
6	Hambantota (two)	14	Saugor Island
7	Trincomalee (two)	15	Chittagong
8	Negapatam	16	Cox's Bazaar

casts were required was on the Burmese coast, the exact location for the first operation was made known only two weeks in advance. This operation, termed DRACULA, was the invasion of Rangoon in the delta of the Irrawaddy. The shortage of time prohibited detailed study of hydrographic data and aerial photographs, as desired. It was evident that forecasting for this region would be particularly difficult because tidal streams reached velocities as high as 7 knots, and shallow water extended miles to sea (FIGURE 11). Fortunately, a Scripps report on the effect of tidal streams on waves was available.¹⁰ Detailed study soon indicated that the problem was not solved by any means, for the Scripps study assumed deep water, and did not treat with the loss of energy waves experience in passing through tens of miles of currents continually increasing in speed.

Wave forecasts were particularly desired in the assembly area offshore, where there was approximately a three knot current and the mean depth of water was about thirty-eight feet. The correction technique suggested by the Scripps report on tidal currents, when augmented by the correction for change in height due to shallow water, appears to give values for wave height that compare favorably with those actually experienced at the site. Sample values obtained by this method are given in Table 3.

Considering the vastness of the area, the meteorological and wave observation network for DRACULA was unusually good. A submarine, HMS STRONGBOW, was stationed just west of the northern tip of Sumatra, and the Second Weather Reconnaissance Squadron, U.S.A.A.F., arranged a special series of flights, termed the DEVIL missions, to the Andaman Sea (FIGURE 10). To insure high caliber wave observations, Lochner accompanied each of these flights as a wave observer and spent over one hundred hours in the air during the ten-day period. Hourly wave observations were radioed with the hourly in-flight weather reports to Calcutta and thence to Colombo. Although made from fast medium bombers (B-25s), the wave observations were generally of high caliber. Wind wave and swell heights were estimated while flying from 50 to 100 feet above the water, and wave period and direction of travel in deep water were generally obtained between altitudes of 1,000 to 4,000 feet. Along coastlines, period and direction were obtained through close scrutiny of breaker zones at lower altitudes. To make certain that undetected, low, long-period swell was not entering the area, the flights swung over Preparis Island, where surf on variously oriented beaches could be readily studied.

At the time of the operation, very low southwesterly swell from a broad, shallow low pressure area in the region of the Andaman Islands was present in the assembly area and combined with waves 2 to 4 feet high, formed by local winds. Low southerly swell from a tropical storm at about 10 degrees, south latitude, also arrived during the operation but was of little operational significance. The official wave forecast for the assembly area, as prepared by the Section, stressed that waves about 4 feet in height would be decreased

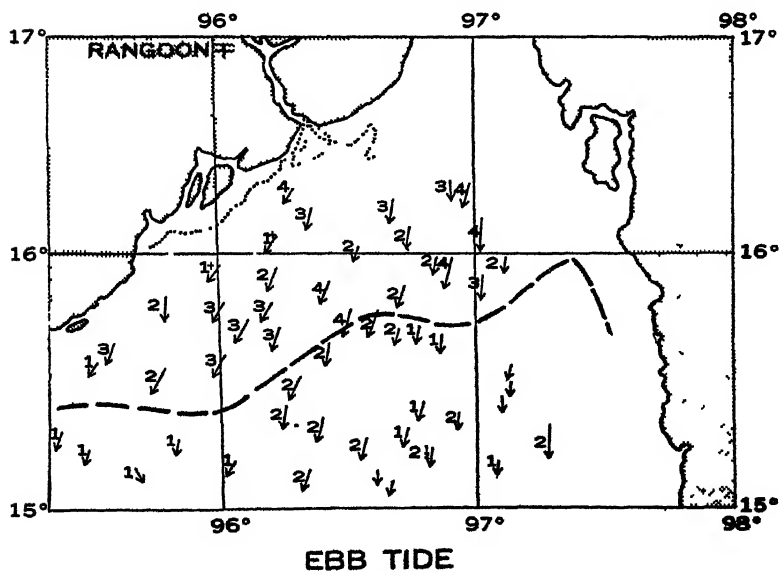
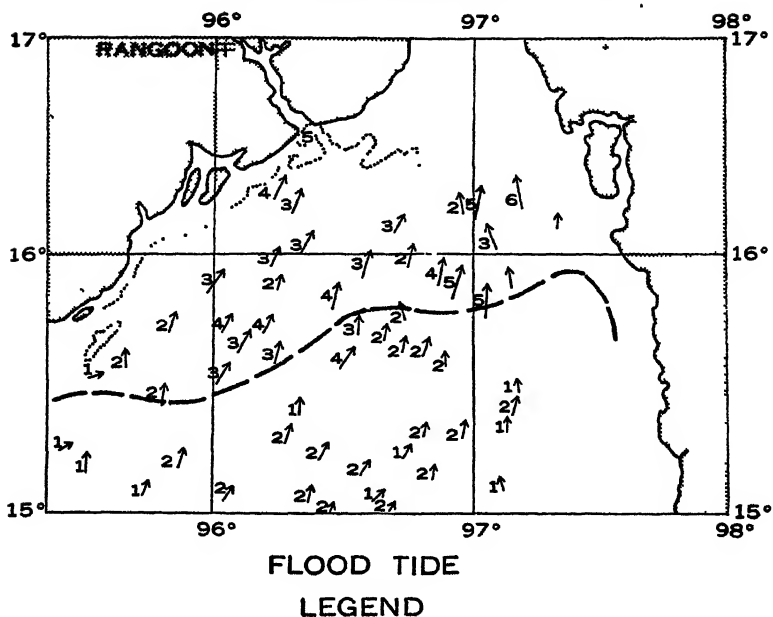


FIGURE 11. Sketch map showing tidal currents in the Gulf of Martaban during ebb and flood tide.

by half a foot on flood tide and increased by 1 to 2 feet on ebb tide. On-the-spot wave forecasts were supplied the invasion fleet by Cauthery, embarked aboard the HMS BLACKMORE, one of the participating destroyers. In keeping with the rule, unusually severe weather conditions occurred just at the time a major operation was scheduled, as in the case of the landings on Sicily, Japan, Leyte, and Normandy. The tropical low in the Andaman area threatened to develop into a storm just at the time of the operation. Luckily, the storm formed only a small but very intense center, about 50 miles in radius, which passed west of the Rangoon area and eventually turned inland near Ramree Island without damaging naval craft.

While the forecasts of the Swell Section had been very satisfactory during the operation, there was no post-assault requirement for such work. The group was free to turn full attention to the problems raised by the projected invasion of Malaya scheduled for September 9, 1945. Timing of this

TABLE 3
PROPOSED WAVE HEIGHTS FOR ASSEMBLY AREA OF DRACULA

Deep water wave height (feet)	Flood tide (3 knot following current)		Ebb tide (2.5 knot opposing current)	
	Period of 6 seconds	Period of 8 seconds	Period of 6 seconds	Period of 8 seconds
2	1.5 feet	1.5 feet	2.5 feet	2.5 feet
4	3	3	5	5
6	4.5	4.5	8	7.5
8	6	6	10.5	10
10	7.5	7.5	13	12.5
12	8.5	9	15.5	15
14	10	10.5	18	17.5

operation was particularly critical since, it would have used both parachute and air-borne troops, as well as troops transported by landing craft. Because of this, Southeast Asia Command was particularly interested in the odds of occurrence when conditions were favorable for landing craft to cross the Bay of Bengal during the monsoon. Because the "hindcasting" technique of establishing sea and swell conditions from past weather maps is not readily applicable to this region, it was decided to establish operational conditions in terms which could be defined by data available in wind, sea, and swell roses for five-degree squares. The problem was far from simple, as sea and swell observations are not highly reliable, being predominantly descriptive in nature, rather than quantitative. In addition, climatologists have usually over-simplified the roses. In so doing, they have often chosen climatic breakdowns that included the entire range from favorable to unfavorable in one category as in the case of wind forces 4 to 7. Determination of operational limits of the landing craft in terms of the natural environment

was also difficult, but values were eventually assigned that did not arouse a storm of criticism. By averaging favorable wind conditions with favorable sea conditions, a fairly reliable value was obtained. This was then corrected for unfavorable swell conditions which occurred when the first two conditions were satisfactory. Computations of this type provided a value for operational conditions for each five-degree square by months, so that one was able to quantitatively compare locations or months, even though individual values by themselves contain appreciable error because of crudeness of

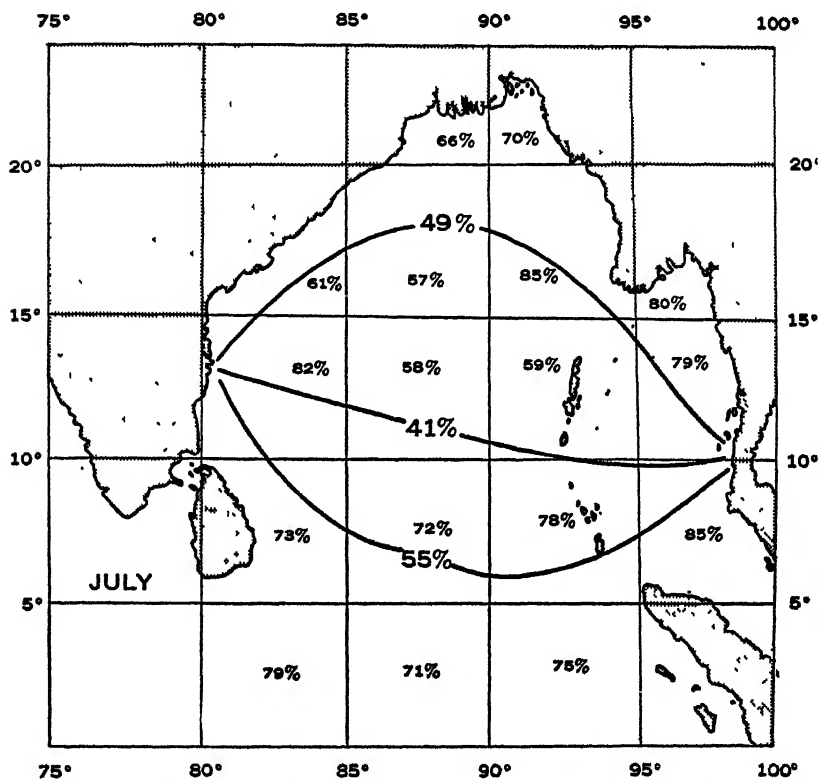


FIGURE 12. Probable frequency of suitable conditions for crossing the Bay of Bengal during the southwest monsoon, as well as suitable conditions for each 5° "square" (month of July).

basic data (see FIGURE 12). Routes could be compared by determining the lowest and highest values of possible occurrence of acceptable conditions along each route and assuming that the true value of acceptable conditions in each case lies at some unknown point between the two extremes. The best estimate is found by taking the mean of the extremes and comparing the values obtained. A technique similar to the one just described for routes was also developed to determine probable frequency of various surf heights along a long strip of simplified coastline.¹¹

Wave Forecasting for the Projected Invasion of Japan

During May, 1945, it became evident that large scale amphibious operations involving Army troops in the western Pacific would require an oceanographic service. To meet this requirement, American personnel of the Swell Section were transferred to China and thence to the Philippines to form the nucleus of the Oceanographic Section of the Far East Air Forces (FEAF) Weather Central at Manila. Once chains of command became clarified, it was determined that the Manila weather group held final responsibility for over-all planning of the Army's invasion weather service. The Oceanographic Section was then instructed to proceed with planning and implementation of a wave forecasting service for operation OLYMPIC, the invasion of Kyushu.

Inasmuch as this, the largest amphibious assault in history, would be undertaken during winter months over three open beaches with varying orientations, wave conditions were bound to be an important factor (see FIGURE 13). To reduce this factor, there were available a back-log of experience in the subject, experienced forecasters in both Navy and Army, and perhaps most important, a realization by the field forces that oceanographic forecasts, if properly used, were of definite value in combating this factor. Since the Navy had primary responsibility for "staging" and "initial assault" phases and the Army primary responsibility during "build-up" and "supply" phases, it was only logical to pool the efforts of meteorologists from both services to make certain that there would be continuity and confidence in the weather service offered to the assault forces. The coordination was relatively easy in the Manila area because FEAF and 7th Fleet Weather Centrals shared the same building and because aerologists to Commander, Amphibious Forces, U. S. Pacific Fleet and to the 3rd and 7th Amphibious Forces were aboard flagships anchored in Manila Bay. Visual wave observations scheduled to begin during September would also have been a joint affair. Weather observers at various points on Okinawa, Iwo Jima, Ie Shima, Guam, Saipan, Samar, and Luzon and aboard four weather ships stationed east and northeast of the Philippines would have reported daily at 0700, 1200, and 1600 hours local time in the Combined Surf Code. As sea conditions could be inferred from wind data, the weather ships would have reported only swell height and direction according to the code, swell taking the place of surf in this instance. Observations would then have been radioed to Manila for inclusion in the routine synoptic broadcast. Installation of wave measuring instruments was not planned, because of shortage of time and equipment.

Meanwhile, as plans for OLYMPIC developed, forecasts were prepared for the Transportation Section of Philippine Base Section, operators of inter-island shipping. The frequent hurricanes of the summer of 1945 soon tested the forecasting techniques developed by Commander W. J. Francis,

U.S.N.,¹² and it became distressingly evident that wind forecasts being made in conjunction with these storms were inadequate to serve as the basis for accurate sea, swell, and surf forecasting. One example of this was the wave forecasts prepared when typhoon "Queenie" crossed northern Luzon from east to west as a weak low during the night of August 5th and reached maximum intensity while about 300 miles west of the island on August 7th.

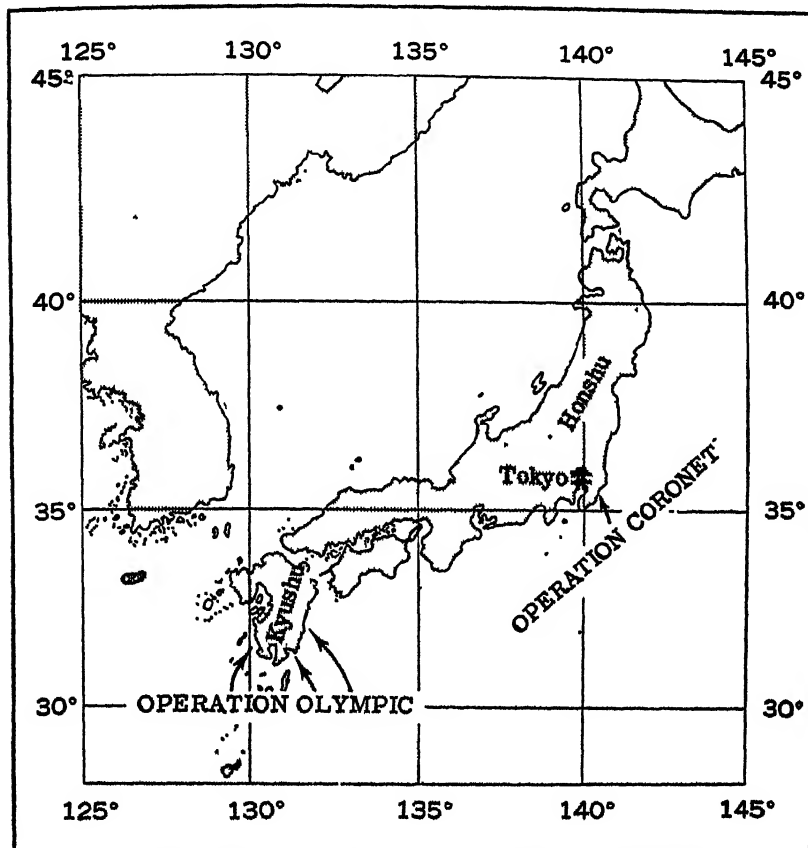


FIGURE 13. Index map showing proposed amphibious assaults against Japan.

A flight by a Navy hurricane reconnaissance plane indicated waves to be about 40 feet high and surface winds in excess of 80 knots in the southern quadrant, the sector from which swell would be generated for the north-western coast of Luzon.¹³ Both Army and Navy forecasts called for breakers to be twelve to eighteen feet high along this coast during the 8th, yet when Lochner flew north during the morning hours, he found breakers to be but two to three feet high! Surface wave observations made just

south of the mouth of the Laoag River during early afternoon gave an average height of three to four feet and a maximum of six feet or a trifle higher. The wave period, if determined from the number of waves in a five minute interval, was six seconds; if determined from a series of sharply defined waves, seven seconds. The storm continued westward to the coast of French Indo-China and wave conditions in the Luzon area soon returned to normal. The sharp discrepancy between forecast and observed wave heights east of this storm can only be explained by a wind distribution entirely different than that expected from the circular isobaric pattern customarily drawn for typhoons. Winds observed by the reconnaissance flight and by surface ships in the southern quadrant within 150 miles of the storm's center bear out this assumption. Of thirteen reports available, only one indicates a westerly wind, while the others range from south-south-east to southwest, directions which would provide a generating area so oriented that swell would pass to the north of the Philippines. Unfortunately, this anomalous situation was not evident at the time, and the rule that wave forecasts are no better than the wind analyses on which they are based was verified with a vengeance.

By the time hostilities ended in mid-August, 1945, the wave forecasting service for OLYMPIC had become well defined. The basic sea, swell, and surf forecasts required by General MacArthur's headquarters and the forces afloat would be supplied by a joint oceanographic section in the weather central building at Manila. This group would also act as the central collecting and disseminating agency for surface and aerial wave observations. An Army oceanographic meteorologist would have been aboard each of the task force flagships to assist the force aerologist and to provide wave forecasts required by the staff Army meteorologist. Once beachheads had been secured, mobile weather stations staffed by oceanographic meteorologists would be established at each beach command to provide weather and wave information in a manner similar to that finally developed in France.

As soon as the beachhead stations were functioning, the army meteorologists aboard the flagships would have been withdrawn to serve as the planning unit for the service needed in operation CORONET, the invasion of the Kanto plain in eastern Honshu. This, the last scheduled large scale landing of the war, would also have been particularly susceptible to wave action, for the beach was directly exposed to oceanic conditions and a shallow bar offshore would frequently have caused an additional surf zone certain to have made difficult the ship-to-shore movement of cargo.

Conclusion

Forecasting of sea, swell, and surf conditions in quantitative terms developed during the early part of World War II and reached maturity within three short years. The techniques were basically correct and could be modified by meteorologists trained in the methods to provide reliable

forecasts for amphibious operations wherever they might be held. The value of the information is well expressed in one of the citations given for wave forecasting in the Normandy invasion. The citation reads, in part, "The work has aided materially in the success of the assault operation and in operations on the beach after the assault. The efforts . . . have been a real contribution to the success of the present campaign."

References

1. MARSHALL, G. C., E. KING, & H. ARNOLD. 1947. The War Reports. J. B. Lippincott. Philadelphia.
2. CORNISH, V. 1934. Waves of the Sea and Other Water Waves. Cambridge University Press. Cambridge, England.
3. SVERDRUP, H. U., & W. H. MUNK. 1946. Empirical and theoretical relations between wind, sea, and swell. Trans. A. G. U. 27: 823-827.
4. HYDROGRAPHIC OFFICE, U. S. NAVY. 1943. Wind waves and swell: principles in forecasting. Navy Department, Washington, D. C.
5. HYDROGRAPHIC OFFICE, U. S. NAVY. 1944. Breakers and surf: principles in forecasting. Navy Department, Washington, D. C.
6. JOHNSON, J. W., M. P. O'BRIEN, & J. D. ISAACS. 1948. Graphical construction of wave refraction diagrams. H. O., Navy Department, Washington, D. C.
7. STEVENSON, T. 1886. The design and construction of harbours. A. & C. Black. Edinburgh.
8. DAVIES, A. 1946. Geographical factors in the invasion and battle of Normandy. Geog. Rev. 36(4): 627.
9. SEIWELL, H. R. 1947. Military oceanography in World War II. Mil. Eng. 39(259): 202-210.
10. SCRIPPS INSTITUTION OF OCEANOGRAPHY. 1944. On wave heights in straits and sounds where incoming waves meet a strong tidal current. Wave Report No. 11 (mimeographed). La Jolla, California.
11. BATES, C. C. 1946. Utilization of tabulated sea and swell data. (Unpublished manuscript).
12. FRANCIS, W. J. 1944. Waves and swell from a tropical storm. Wave Report No. 29 (mimeographed) of Scripps Institution of Oceanography. La Jolla, California.
13. CHIEF OF NAVAL OPERATIONS, U. S. NAVY. 1945. Typhoon reconnaissance, June through September, 1945. Navy Department, Washington, D. C.

Discussion of the Paper

WARREN C. THOMPSON (*Scripps Institution of Oceanography, La Jolla, California*): An unpublicized phase of sea, swell, and surf forecasting was developed during the war by the Navy Amphibious Forces in the Pacific. This work centered around the Staff Aerological Officer and a team of flying aerologists trained in both forecasting and observing wave and beach conditions on enemy beaches. The team was formed with the object of avoiding needless losses of men and equipment that might be caused by adverse hydrographic conditions in the landing phase of an invasion, such as occurred at Tarawa. The hydrographic information obtained over enemy beaches was used by amphibious commanders in making decisions as to the desirability of utilizing the designated landing beaches. It also served as a guide for boat captains in the initial landing through the surf.

The three members of this unique reconnaissance team were Lieutenant W. M. Johnson, U.S.N.R., Lieutenant D. B. Murphy, U.S.N.R., and Lieutenant W. C. Thompson, U.S.N.R. They began their careers as observers when they enrolled in the Navy course in sea and swell forecasting

which was set up under the guidance of Captain H. T. Orville, U.S.N., at Scripps Institution of Oceanography. The course was organized by Dr. H. U. Sverdrup, Director of the Institution, who, with the cooperation of Mr. W. H. Munk, developed the practical technique of wave forecasting which was later to be used by the observers in predicting wave conditions for invasions.

In August, 1944, the three officers were assigned to the staff of the Third Amphibious Force, commanded by Vice Admiral T. S. Wilkinson, U.S.N. The Hydrographic and Special Beach Observers, as they were unofficially designated, were under the immediate supervision of Commander O. D. Finnigan, U.S.N., Staff Aerologist, and were extensively trained by him. Their job was essentially to make pre-invasion aerial observations of weather, sea, and surf conditions at a given beach, extent and type of obstacles in the approaches to the beach, gradient and width of the beach, and any other pertinent data. In order to discharge these duties the observers were required to become proficient in aerial observation, beach hydrography and intelligence, photo interpretation, and aircraft communications. In addition, they were required to have a broad knowledge of amphibious operations, and a detailed knowledge of swell and surf conditions and their effect on amphibious craft and vehicles.

Trial aerial observations were made on the beaches of Oahu and it was found after a little practice that the breaker heights that were observed up to 6-8 feet could be estimated quite consistently within a foot of the actual height when flying at altitudes of 1000 to 2000 feet. The estimated heights were verified by observers on shore. This accuracy of visual estimation fixed confidence in the plan of making surf observations from the air.

The first operation assigned to the Third Amphibious Force, after succeeding the Fifth Amphibious Force in the Pacific offensive, involved the seizure of the southern Palau Islands in order to provide airfields for the forthcoming Philippine invasion. Thompson was chosen to make the first beach reconnaissance and joined the advance force at Guadalcanal. He was necessarily billeted aboard an aircraft carrier and became a member of the flight crew as a Technical Observer. Upon arrival at Peleliu, he made frequent pre-D-day familiarization flights in TBM-type aircraft over the landing beaches.

While aboard the carrier off Peleliu, a continuous forecast of sea and swell was maintained to determine the wave conditions that should be expected in deep water adjacent to the barrier reef which fringed the landing beaches. Since no method of forecasting wave heights and breakers over a submerged reef had been devised, the swell forecasts were valuable only to give an idea of the magnitude of the surf activity. The forecasting procedure used was that which was developed by Sverdrup and Munk, and it proved to be quite accurate and adequate for its purpose.

Lieutenant Commander G. H. Heyen of the Royal Australian Naval

Reserve, on special duty with the U. S. Navy, also acted as a hydrographic observer of the Peleliu beaches. Heyen was engaged because of his extensive knowledge of tropical weather and sea conditions acquired through many years of living and sailing among the coral islands. Before dawn on D-day, 15 September 1944, the two observers flew over the Peleliu landing beaches until it became light enough to see. Comprehensive observations of the surf conditions were then made as quickly as possible, because time as well as accuracy was essential, and the information was relayed by radio to the flagship of the operation. Three or more passes over a beach were required to collect the necessary information, which consisted of breaker heights and periods, type of breaker, width of the surf zone, direction of sea and swell, and so forth. Flights were usually made at altitudes of 1000 to 2000 feet, but at Peleliu they were made as low as 20 to 50 feet. The Palau landing began about three hours after the hydrographic observations were completed. The surf conditions for landing were very good and breaker heights did not exceed two feet. After the landing the main job of the observers was completed. Subsequent observations for invasions were obtained by a similar procedure.

Preliminary invasion plans called for an assault on Yap Island, and Johnson and Murphy were assigned to cover the beaches. However, with stiff opposition encountered at Peleliu and a weaker enemy air strength than anticipated in the Philippines, the Yap operation was canceled. The Yap expedition was diverted to the Philippines and was augmented for a landing on Leyte, which involved both the Third and Seventh Amphibious Forces under the overall command of Vice Admiral Kinkaid, U.S.N., Commander Seventh Fleet. Aerial observations of beach and surf conditions were made at Leyte by Johnson and Murphy several days prior to and during landing operations on 20 October 1944.

Johnson and Murphy next made observations for the landings in Lingayen Gulf. In connection with their sea and swell forecasts, they prepared refraction diagrams for the Gulf, and, on the basis of the diagrams, they made recommendations for landing men and supplies on stretches of beach which displayed the greatest divergence of wave orthogonals. Their advice was well borne out by adverse sea conditions in the Gulf following the operation, and for this timely piece of work Johnson and Murphy were awarded the Commendation Ribbon.

Following the Philippine landings, the Third Amphibious Force retired and the Fifth Amphibious Force resumed the offensive. It opened with a landing on Iwo Jima on 19 February 1945. Prior to and including D-day, Johnson and Murphy reconnoitered the assault beaches on the west side of the island and Thompson covered the alternate beaches on the east side. Whenever it was possible, alternate beaches were chosen for landings so that, if weather and sea conditions on the main assault beaches made landing inadvisable, the alternate beaches might be available for attack. The

aerial observations at Iwo Jima were somewhat hampered by cloudy weather, but the sea conditions were found to be very good for the landing on D-day. Several days after, however, high surf conditions which were forecasted temporarily suspended the landing of vitally needed supplies on the assault beach.

The final set of beach and surf observations were made by the three observers at Okinawa. On D-7 day Thompson began making observations of the Karama Retto, a group of small islands a few miles to the west of southern Okinawa. The Retto was occupied because it forms for ships a natural shelter from adverse sea and weather conditions. A few days prior to D-day, Johnson and Murphy began their reconnaissance of the main landing beaches on the west coast of Okinawa and Thompson covered the alternate beaches on the southern coast. On D-day, 1 April 1945, conditions for landing were again quite favorable.

The job of aerial beach observer was not without its hazardous moments, aboard ship as well as in the air. Johnson experienced his first real action aboard the USS FANSHAW BAY when she was heavily shelled by enemy surface craft in the Philippine Sea. Also in the Philippines, Murphy was aboard the USS KITKUN BAY when it was badly damaged by a suicide plane. Later, Thompson was aboard the USS WAKE ISLAND off Okinawa when she also was damaged by a suicide plane. Thus, for obvious reasons the three observers were not generally together aboard the same ship during an operation. However, on D + 2 day off Iwo Jima, they gathered aboard the USS BISMARCK SEA to take part in a fast carrier strike at Okinawa. As fate would have it, two suicide planes attacked the ship at dusk, and a short time later she slipped beneath the waves. It was fortunate that all three observers were rescued, because there were no trained personnel to replace them at that time.

Toward the end of the Pacific hostilities, three additional aerial observers were being trained, but with the Japanese capitulation on 14 August 1945 the landing beach studies were terminated. As a result of their flight activities, the three observers were presented with the Air Medal with Gold Star by Vice Admiral Wilkinson, U.S.N., in a Navy Day ceremony (1945) in Yokohama. At a later date, each was presented separately with the Bronze Star Medal for "heroic and meritorious achievements in the carrying out of their assignments as Hydrographic and Special Beach Observers."

ANNALS OF THE NEW YORK ACADEMY OF SCIENCES

VOLUME 51, ART. 4. PAGES 575-852

May 31, 1949

MOLECULAR INTERACTION*

Consulting Editor: R. M. FUOSS

CONTENTS

	PAGE
Light Scattering in Soap Solutions. <i>By</i> P. DEBYE.....	575
The Role of Hydration in the Debye-Hückel Theory. <i>By</i> R. A. ROBINSON AND R. H. STOKES.....	593
Some New Procedures in Thermodynamic Theory Inspired by the Recent Work of J. N. BRØNSTED. <i>By</i> VICTOR K. LAMER, OLAV FOSS, AND HOWARD REISS.....	605
The Effects of Shape on the Interaction of Colloidal Particles. <i>By</i> LARS ONSAGER..	627
The Attractions of Proteins for Small Molecules and Ions. <i>By</i> GEORGE SCATCHARD.	660
Determination of the Dissociation Constants of Weak Electrolytes in Salt Solutions. <i>By</i> MARTIN KILPATRICK.....	673
Thermal Properties of Electrolytic Solutions and the Debye-Hückel Theory. <i>By</i> FRANK T. GUCKER, JR.....	680
Interactions Between Vapors and High Polymers. <i>By</i> MALCOLM DOLE.....	705
The Scientific Contributions of Grinnell Jones. <i>By</i> MALCOLM DOLE.....	719
Pressure-Volume-Temperature Relations in Solutions VIII. The Behavior of Some Solutions of Electrolytes in Water. <i>By</i> R. E. GIBSON AND O. H. LOEFFLER....	727
Extrapolation of Apparent Molal Properties of Strong Electrolytes. <i>By</i> BENTON B. OWEN AND STUART R. BRINKLEY, JR.....	753
The Surface Tensions of Aqueous Sulfuric Acid Solutions. <i>By</i> T. F. YOUNG AND S. R. GRINSTEAD.....	765
The Diffusion Coefficient of Potassium Chloride in Aqueous Solution at 25°C. <i>By</i> HERBERT S. HARNED AND RALPH L. NUTTALL.....	781
Some Interactions in Solutions of Electrolytes. <i>By</i> C. A. KRAUS.....	789
Dielectric Absorption by Ammonium Salts in Solvents of Low Dielectric Constant. <i>By</i> ROBERT H. COLE AND HOWARD A. STROBEL.....	807
Factors Influencing the Properties of Long-Chain Electrolytes in Water and Water Mixtures of Organic Solvents. <i>By</i> PHILIP F. GRIEGER.....	827
The Viscosity of Mixtures of Polyelectrolytes and Simple Electrolytes. <i>By</i> RAYMOND M. FUOSS AND ULRICH P. STRAUSS.....	836

*This series of papers is the result of a Conference on Molecular Interaction, held by the Section of Physics and Chemistry of The New York Academy of Sciences on April 9 and 10, 1948.
Publication made possible by a contribution from the General Funds of The Academy.

Copyright, 1949, by

The New York Academy of Sciences

LIGHT SCATTERING IN SOAP SOLUTIONS

BY P. DEBYE*

Baker Laboratory, Cornell University, Ithaca, New York

Introduction

Soap solutions exhibit even lower osmotic activity than would be predicted if one assumed that soap existed in solution as simple undissociated molecules. Soap solutions also conduct the electric current far better than would be expected from the observed osmotic effects. Attempting to explain these anomalies, McBain,¹ in 1913, suggested that the fatty soap ions aggregated in solution. Such colloidal aggregations of ions, which were termed micelles, would explain the low osmotic activity and relatively high conductivity of soap solutions.

Since 1913, investigators have shown considerable interest in the determination of the size and shape of the micelle. McBain² proposed two different micelle species, which he said could coexist in solution: one a small, spherical, hydrated, ionic micelle, and the other a large, lamellar, weakly conducting micelle. While agreeing with McBain that the behavior of soap solutions pointed to the existence of micelles, Hartley³ took the view that only the small spherical micelle was feasible. On the basis of geometrical considerations, Hartley⁴ calculated that the micelle of a 16 carbon soap consisted of approximately 50 cetyl chains. He and Runnicles⁵ carried out diffusion experiments with cetyl pyridinium chloride and calculated from their results that the micelle of this soap contained about 70 paraffin chains.

Ultracentrifuge and diffusion measurements by Miller and Andersson⁶ on Duponal (sodium salts of sulfated aliphatic alcohols of chain length C_8 to C_{18}) led to a molecular weight of 12,500 for the mixed micelle.

Hakala⁷ has made diffusion measurements on sodium dodecyl sulfate solutions. If a spherical model for the micelle is assumed, the introduction of his results into the Stokes-Einstein equation gives a value of 23.6 Å for the radius. A molecular weight of about 25,000 (87 paraffin chains per micelle) is obtained from this value of the radius if a density equal to that of dodecane is taken.

Vetter⁸ studied the sodium salt of sulfonated di (2-hexyl) succinate, known commercially as Aerosol MA, and from density, viscosity, and diffusion data calculated an aggregation number of 24 for the micelle.

Gonick and McBain⁹ obtained cryoscopic evidence of micelle formation in aqueous solutions of several non-ionic detergents. Assuming ideal behavior, their data indicate that a micelle consists of no more than 7 detergent molecules.

* All the measurements quoted in this paper have been made by E. W. Anacker in the Chemistry Department of Cornell University. His work was supported by the Office of Rubber Reserve.

X-ray work of various German investigators has been cited by McBain¹⁰ as proof of the presence of lamellar micelles in solution. The more recent X-ray work of Harkins¹¹ and co-workers at the University of Chicago tends to disprove the lamellar micelle theory. Their results may be interpreted as showing that the micelle consists of a double layer of soap molecules, roughly cylindrical in shape. According to this model, the hydrocarbon chains in each layer are in alignment and form the body of the cylinder. The polar groups make up the ends of the cylinder. The Chicago group has also calculated from their data the number of soap molecules per micelle. Their values range from 30 molecules per micelle for a 9 per cent by weight potassium laurate solution to 270 molecules per micelle for a 14.1 per cent solution of dodecylamine hydrochloride.

It should be stated, at this point, that the diffusion and ultracentrifuge investigations mentioned were carried out in aqueous salt solutions of the soaps. Hakala, Hartley and Runnicles, and Vetter found that, above a certain salt concentration, the diffusion coefficients were constant and independent of salt or soap concentration. According to Vetter,⁸ "This behavior indicates constancy of micelle size, micelle shape, and degree of solvation. . . if these three factors do not change simultaneously in such a manner that they neutralize their separate effects. The latter possibility seems highly improbable." The work of the Chicago group indicates that the micelle increases in size with increasing soap concentration. Light scattering work seems to indicate that the size of the micelle is constant in a given solvent in at least a range of soap concentrations above the so-called "critical concentration" for micelle formation. It also appears that the size of the micelle is dependent upon the salt concentration of the solvent.

It would not be correct to say that the results from the various methods are in complete disagreement, since all of the investigations mentioned were conducted under different conditions and, for the most part, with different soaps. If the soaps used were the same, the concentrations most likely were not. Soap solutions of at least 9 per cent by weight were used in the X-ray work of the Chicago group, whereas diffusion experiments by Hakala were carried out with solutions whose soap concentrations were about 1.5 per cent or less.

Light scattering offers an independent method of determining the number of solute particles in a given solution and thus enables one to determine the molecular weight of the solute. The quantities which have to be measured for such a determination are turbidities of a series of solutions at different concentrations and the difference in refractive index between the solvent and one solution of known solute concentration. If the particles under investigation have a dimension larger than one-twentieth of the wave length of the light employed, an additional quantity, the dissymmetry coefficient, must be found. This is not necessary for the case in hand.

The turbidity τ is defined as the fractional decrease of the intensity I of the incident beam of light due to scattering per unit length l of its path

$$\tau = -\frac{1}{I} \frac{dI}{dl}. \quad (1)$$

For small particle dimensions, such as are encountered in soap solutions, the angular distribution of the scattered light does not depend on the properties of the particle and a measurement of the intensity scattered at a 90° angle to the incident beam is adequate.

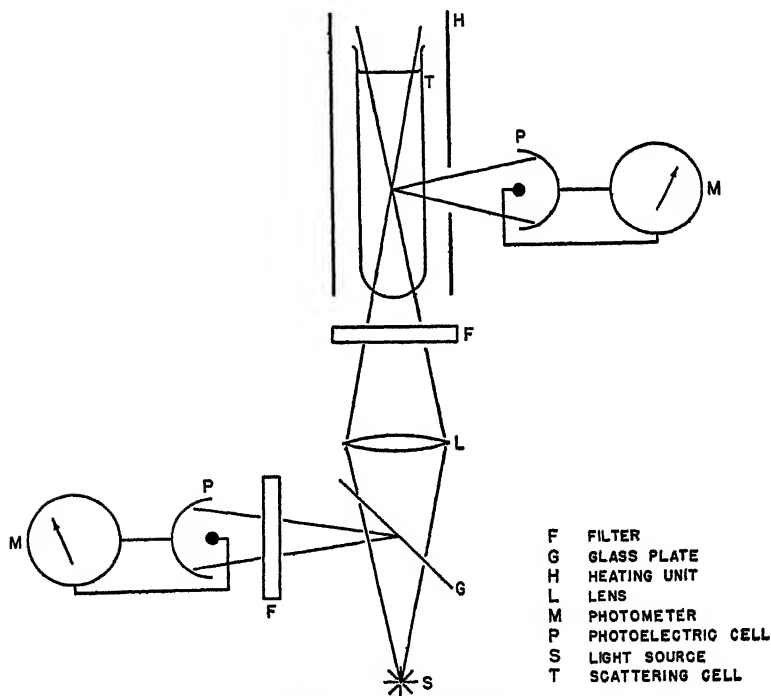


FIGURE 1. 90° scattering instrument

We have used the simple fluorimeter type of instrument illustrated schematically in FIGURE 1.

The light source is a low pressure mercury arc (G.E. AH-4), which is located at the bottom of the instrument. The primary beam passes upward through a lens and monochromatizing filters and is focused in the center of the scattering cell, opposite the photo tube. The deflection of the needle of a current amplifier connected to the photo tube is proportional to the intensity of the scattered light and hence to the combined turbidity of the solution and scattering cell. The proportionality constant is obtained by noting the deflection of the needle of the current amplifier when the standard of known

turbidity is placed in the instrument. The significant result is the excess turbidity of a solution over that of the solvent, which we obtain by subtracting the turbidity of the scattering cell and solution from the turbidity of the scattering cell and solvent.

In order to detect and correct for fluctuations in the primary intensity, a glass plate is set at an angle of 45° in the path of the primary beam. The light reflected by this glass plate is intercepted by a second photoelectric cell. Readings of an amplifier connected to this photo tube give the intensity of the primary beam. A voltage stabilizer and ballast lamp are included in the electrical circuit to help stabilize the output of the lamp.

An insulated brass cylinder wound with resistance wire is mounted vertically in the instrument so as to surround the scattering tube when the latter is placed in position. Turbidities at temperatures ranging from that of room to approximately 90°C. may be measured.

Round bottom pyrex test tubes (25×200 mm.) have been used as scattering cells. Experimentally measured excess turbidities are independent of the particular tube used as long as the tube is thoroughly cleaned and oriented in the instrument in the same manner for each measurement.

Two methods can be used to obtain the turbidities of a series of solutions of different solute concentrations. With polymer solutions, one may start with a 50 cc. portion of solvent in the scattering tube and add to this small measured portions of a concentrated polymer solution. Turbidities are measured after each addition. With soap solutions, it has been found necessary to alter this procedure since large concentration increments are often needed to get appreciable changes in the turbidity. We preferred to make up a series of solutions at desired concentrations and to measure the turbidities of each solution separately. This second method takes more time than the first, but has an advantage in that each turbidity measurement is an independent one. In the other method, the introduction of dust or an error in dilution will violate all measurements thereafter.

Needless to say, all dust and foreign material must be removed from the solutions if accurate results are to be obtained. In our work with soap solutions, we have found it convenient and satisfactory to remove dust by filtering the solutions through a fine sintered glass filter. In scattering measurements, it is a good precaution to distill solvents before use. The water used in our studies of soap solutions was freshly distilled.

In general, $\mu - \mu_0$ for 1 per cent polymer or soap solution is of the order 10^{-3} . Since this difference in refractive index between the solution and solvent occurs squared in the refraction constant H , its value must be determined with some degree of accuracy. Ordinary commercial refractometers are not adequate for this purpose. A suitable differential refractometer has been described by P. P. Debye.¹² We have used such an instrument with occasional minor variations of the cell construction.

Interpretation of Light Scattering

The theory of light scattering has been described in greater detail elsewhere.¹³ For randomly distributed, small particles the result of the theory may be described by the equation

$$\tau = \frac{32\pi^3}{3} \frac{\mu_0^2(\mu - \mu_0)^2}{\lambda^4} \frac{1}{n}, \quad (2)$$

where τ denotes the excess turbidity as defined by EQUATION 1, μ is the refractive index of the solution and μ_0 that of the solvent. λ stands for the wave length (in vacuum) of the light and n for the number of particles per cubic centimeter.

Relation (2) may be rewritten in terms of the molecular weight M and the concentration c (in grams per cubic centimeter) as follows

$$\tau/c = HM. \quad (3)$$

Here, the constant of proportionality H must be derived from refraction measurements according to the equation

$$H = \frac{32\pi^3}{3} \frac{\mu_0^2}{N\lambda^4} \left(\frac{\mu - \mu_0}{c} \right)^2 \quad (4)$$

where N denotes Avogadro's number $(6.0228 \pm 0.0011) \times 10^{23}$.

According to EQUATION 2, the light scattering from a solution of colloidal particles should be proportional to the concentration. This relation, however, holds only exactly in the limit for infinite dilution. For most solutions of polymers, the scattering per particle tends to decrease with increasing concentrations. A typical case is illustrated in FIGURE 2, where it is seen that the effect is more pronounced the better the solvent. Now EQUATION 2 depends on two assumptions. One is that the refractive index is a linear function of the concentration. This hypothesis can be tested directly by refraction measurements and, in nearly all cases, the linear relation is found valid to a high degree of approximation. The second hypothesis is, that not only in calculating the refraction, but also in the discussion of the total scattering the particles can be considered to act independently of each other. This last assumption, however, although correct for high dilutions, begins already to fail at rather low concentrations, well within the range in which experiments are performed ordinarily.

In order to discuss this interaction effect, we can go back to a theory developed by Einstein¹⁵ for the additional scattering from a solution as compared to the scattering from the solvent according to which the difference in question is due to the spontaneous fluctuations of the concentration. Their magnitude, in turn, depends on the osmotic properties of the solution. This also implies, of course, that scattering measurements can be used as a sub-

stitute for straightforward osmotic measurements. Einstein's result may be written in the form

$$\tau = \frac{32\pi^3}{3} \frac{\mu_0^2}{N\lambda^4} \frac{(c\partial\mu/\partial c)^2}{c \frac{\partial}{\partial c} \left(\frac{P}{RT} \right)} \quad (5)$$

where P denotes the osmotic pressure and $R = Nk$ is the gas constant. This formula is valid as long as the dimensions of the particles are reasonably small compared to λ . The variation of μ with c may be determined in an

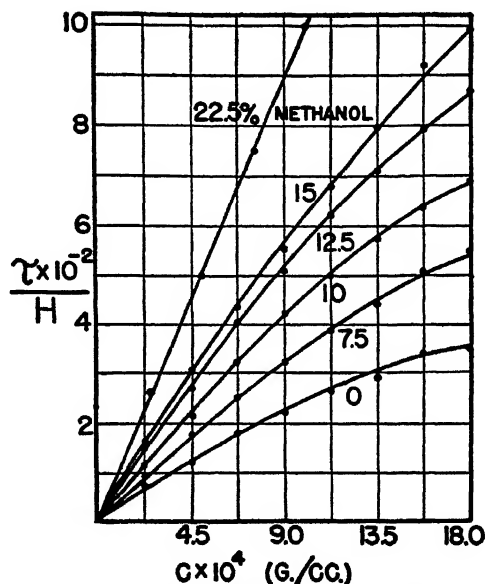


FIGURE 2. Turbidity of polystyrene in benzene-methanol mixtures (McCartney¹⁴).

auxiliary measurement; as mentioned above it is usually linear, so that $\frac{\partial\mu}{\partial c}$ may be replaced by $(\mu - \mu_0)/c$. Thus, we are led to transform EQUATION 5 into

$$H \frac{c}{\tau} = \frac{\partial}{\partial c} \left(\frac{P}{RT} \right) \quad (6)$$

in which H is the same refraction constant as defined by EQUATION 4.

For dilute solutions, according to van't Hoff's law, P/RT is equal to c/M and EQUATION 6 simplifies to EQUATION 3 in this case.

The light scattering measurements illustrated in FIGURE 2 confirm what has been inferred from other experiments: that solutions of high polymers commonly exhibit considerable deviations from van't Hoff's law even in

rather dilute solutions. It has been found that a two-term expression of the form

$$\frac{P}{RT} = \frac{C}{M} + Bc^2 \quad (7)$$

is often valid over a wide range of concentrations. In such cases, we expect, instead of EQUATION 3, the relation

$$H \frac{c}{\tau} = \frac{1}{M} + 2 Bc. \quad (8)$$

In FIGURE 3, in which the experimental results of FIGURE 2 have been used in order to plot Hc/τ as a function of the concentration, it is seen that now we obtain straight lines in accord with EQUATION 8.

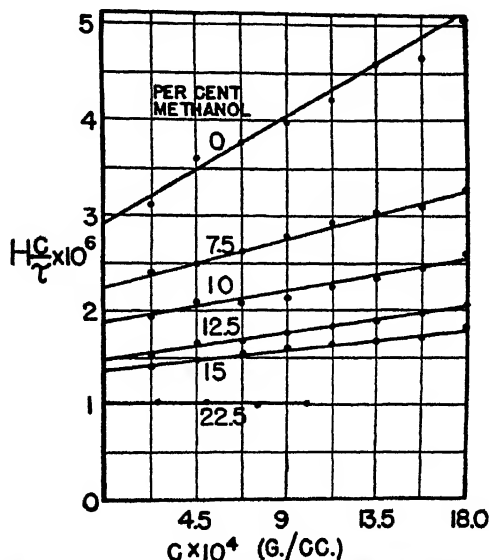


FIGURE 3 Reciprocal specific turbidity of polystyrene in benzene-methanol mixtures.

The intercept at $c = 0$ in representations like FIGURE 3 is equal to $1/M$.

A peculiar feature of the straight lines in our case of mixed solvents is that they have not the same intercept. Considered superficially, this seems to indicate that the molecular weight increases with increasing content of non-solvent in the mixture. It can, however, be shown that the effect is due instead to preferential adsorption of the benzene on the polymer particle¹⁴ and its interpretation furnishes a quantitative measurement of this preference. In simple solvents, no such effect occurs, except for possible real agglomeration of the polymer particles.

Mass Action

The relations valid for high polymers are not directly applicable to soap solutions. However, we shall presently consider the mass action equilibrium between simple ions and micelles, and arrive at an extremely simple modification of these relations.

Consider the following idealized reaction between fatty ions A and micelles A_n , where n is the number of fatty ions per micelle:



If we let c_n be the concentration of micelles, c_1 the concentration of unaggregated paraffin chains, and c the total concentration of fatty ions the following relationships hold:

$$\frac{c_1^n}{c_n} = K \quad (10)$$

$$c = c_1 + nc_n \quad (11)$$

K is the equilibrium constant. It has been assumed that, for this simple treatment, activity coefficients are equal to unity.

The equilibrium constant K has the dimension of a concentration to the power $(n - 1)$. We write $K = c_0^{n-1}$ and express our concentrations as multiples of c_0 . For the relative concentrations

$$\gamma_1 = \frac{c_1}{c_0}, \quad \gamma_n = \frac{cn}{c_0}, \quad \gamma = \frac{c}{c_0}, \quad (12)$$

the relations

$$\gamma_1^n = \gamma_n, \quad \gamma_1 + n\gamma_n = \gamma \quad (13)$$

hold.

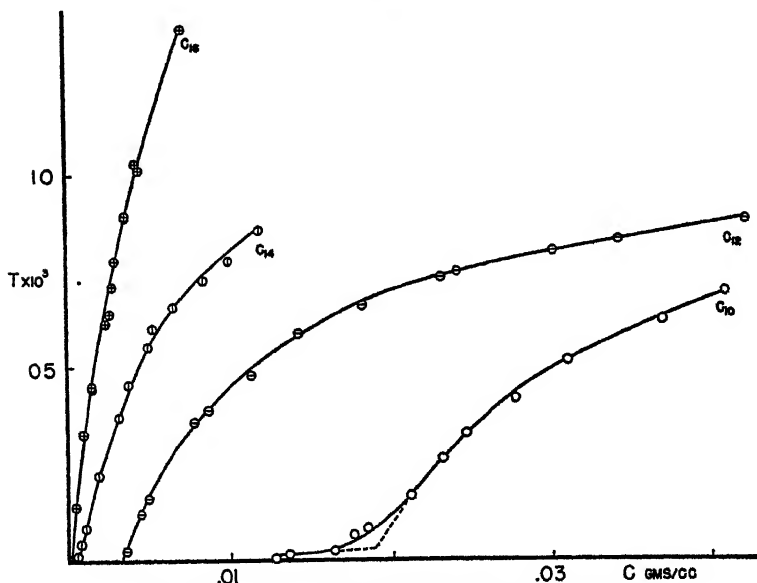
For very large values of n , it turns out that the relative concentration γ_1 of the monomer is equal to γ for $\gamma < 1$. From $\gamma = 1$ on the concentration γ_1 remains constant. The relative concentration of the polymeric particle, on the other hand, is 0 from $\gamma = 0$ to $\gamma = 1$ and equal to $\gamma - 1$ from there on. It is seen that $\gamma = 1$ corresponds to a critical point and we shall have to identify c_0 with the critical concentration.

For large but finite values of n , the sharp edges of the curves for γ_1 and γ_n at $\gamma = 1$ are rounded off. A straight-forward calculation for $n = 65$ leads to the values of TABLE 1. Under these circumstances, there would be some micelles immediately below the critical concentration (see the first part of the third column), whereas immediately above this concentration, more micelles would be present as expected for n infinite (see the fourth column). These considerations indicate that the soap is practically unaggregated up to the critical concentration c_0 and that, from here on, all but a negligible fraction of the soap which is added in excess will appear in the

form of micelles. In FIGURE 4, excess turbidities of solutions of the 10, 12, 14, and 16 carbon *n*-alkyl trimethylammonium bromides above that of the solvent are plotted against total soap concentration *c*. Our qualitative expectations are confirmed. The fact that little or no turbidity difference

TABLE 1

γ	γ_1	$n\gamma_n$	$n\gamma_n - (\gamma - 1)$
0.8408	0.8400	0.0008	—
0.8636	0.8600	0.0036	—
0.8960	0.8800	0.0160	—
0.9234	0.8900	0.0334	—
0.9690	0.9000	0.0690	—
1.0514	0.9100	0.1414	0.0900
1.2079	0.9200	0.2879	0.0800
2.1051	0.9400	1.1651	0.0600
5.5360	0.9600	4.5760	0.0400
18.472	0.9800	17.592	0.0200
34.835	0.9900	33.845	0.0100
66.000	1.0000	65.000	0.0000

FIGURE 4. Turbidities of *n*-alkyl trimethylammonium bromides in 0.0130M KBr.

between solutions and solvent could be detected means that either no micelles exist at low concentrations or that their number is very small. The rapid rise of the curve for concentrations just above the critical concentration indicates that here the micelles are rapidly increasing in number. From the critical concentration on, the curves resemble typical τ vs. c plots for polymer

solutions. This then means that, in treating the light scattering data for soap solutions, we should not plot Hc/τ but $H\frac{(c - c_0)}{\tau}$ as a function of the total soap concentration c ; and that we should extrapolate to $c = c_0$ rather than to $c = 0$ in order to obtain the reciprocal of the molecular weight.

In FIGURE 5, data for aqueous solutions of the 10, 12, and 14 carbon trimethylammonium bromides are plotted in accordance with the preceding discussion. The heights of the vertical lines drawn at the critical concentrations represent the reciprocal molecular weights of the micelles. It is readily seen that the molecular weight increases as the hydrocarbon tail of the soap molecule is lengthened.

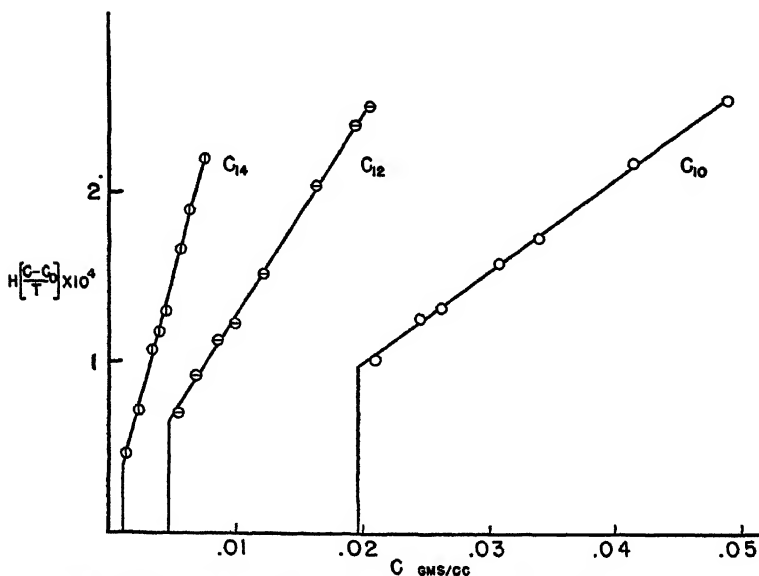


FIGURE 5. Reciprocal specific turbidities of *n*-alkyl trimethylammonium bromides in water.

In FIGURE 6, are plotted the observed data for aqueous and salt solutions of dodecylamine hydrochloride. We see that the addition of salt lowers the critical concentration and increases the size of the micelle. The decrease in slope shows that the solvent becomes progressively more poor as salt is added. If enough salt is added, the soap will be "salted out." The molecular weights (TABLE 2) are plotted against the added chloride ion concentration in FIGURE 7. Although the single run made with $BaCl_2$ solution as the solvent is not sufficient to draw definite conclusions, the results of that run indicate that the size of the micelle is dependent only upon the concentration of the ion opposite in charge to that on the fatty ions comprising the micelle.

The critical concentrations of dodecylamine hydrochloride in the various salt solutions were taken from a curve plotted from data given by Corrin

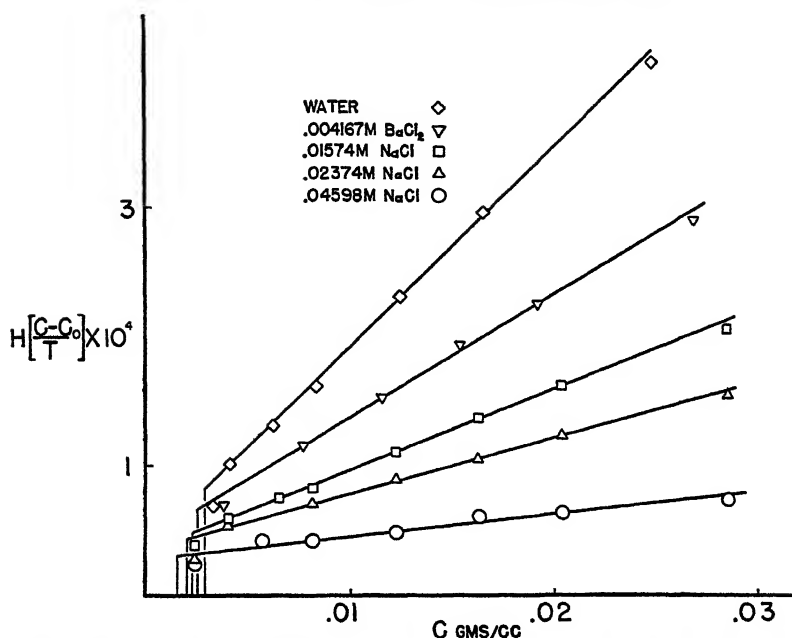


FIGURE 6. Reciprocal specific turbidity of dodecylamine hydrochloride in water and salt solutions.

TABLE 2
DODECYLAMINE HYDROCHLORIDE IN AQUEOUS AND SALT SOLUTIONS

Solvent	Crit. conc.	M.W.
Water	$1.31 \times 10^{-3} M$	12,300
.004167 M BaCl ₂	$1.14 \times 10^{-3} M$	14,600
.01574 M NaCl	$1.04 \times 10^{-3} M$	20,500
.02374 M NaCl	$9.25 \times 10^{-4} M$	22,400
.04598 M NaCl	$7.22 \times 10^{-4} M$	31,400

and Harkins.¹⁶ All other critical concentrations given hereafter have been determined by a least squares method, which also gives at the same time the best linear plots of $H \frac{(c - c_0)}{\tau}$ vs. c .

In FIGURE 8 are $H \frac{(c - c_0)}{\tau}$ vs. c plots for the n -alkyl trimethyl bromides in 0.0130 M . KBr. From a comparison of this figure with FIGURE 5 and the results listed in TABLE 3, we see that the effect of salt is much more pronounced for the longer chain lengths than for the short. The n -decyl soap micelle is scarcely changed in size, while the micelle of the n -tetradecyl micelle increases in size by a comparatively large amount.

If we consider the change made in the "total" bromide ion concentration (TBIC) when bromide ion is added in the form of a salt, the resulting changes

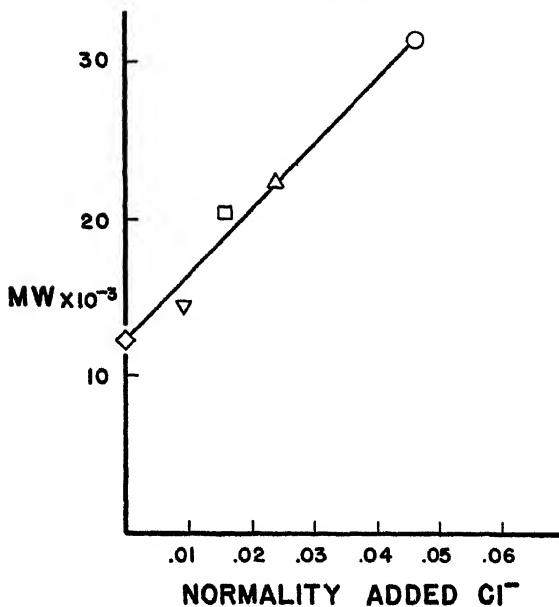


FIGURE 7. Molecular weight of dodecylamine hydrochloride as a function of added Cl^- .

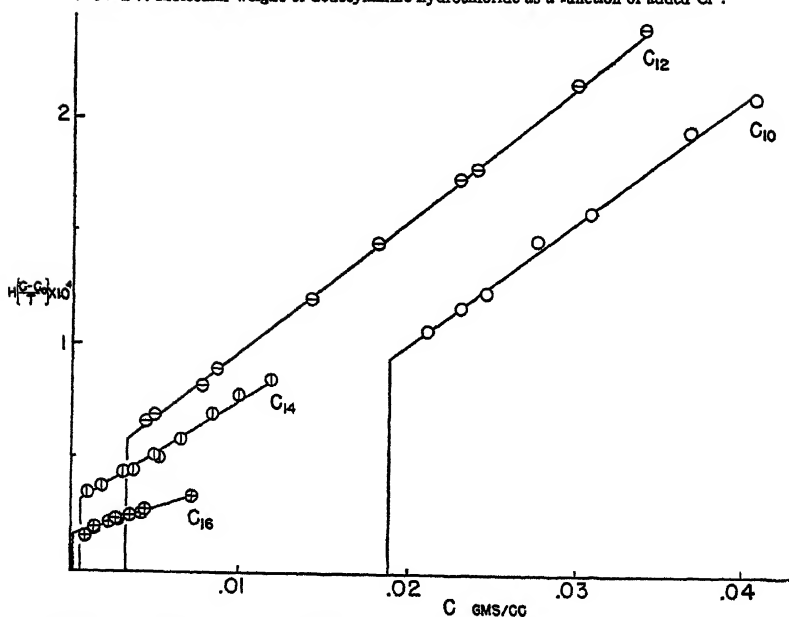


FIGURE 8. Reciprocal specific turbidities of n -alkyl trimethylammonium bromide in 0.0130M KBr.

in micelle size for the various soaps are not startling. By TBIC, we mean the sum of the critical concentration and the concentration of the added

bromide ion. We are tacitly assuming that the charge of the micelle is zero. To emphasize this point, we might compare the results obtained for C_{10} and C_{14} in water and 0.0130M KBr. In the former case, we increase TBIC from 0.0700 to 0.0802 and N from 36 to 38. In the latter case, we increase

TABLE 3
n-ALKYL TRIMETHYLAMMONIUM BROMIDES IN WATER AND 0.0130M KBr

Solvent	Soap	C_0	MW	Monomeric soap ions per micelle
Water	C_{10}	0.0700M	10,200	36
	C_{12}	0.0151	15,500	50
	C_{14}	0.00341	25,300	75
0.0130M KBr	C_{10}	0.0672	10,700	38
	C_{12}	0.0105	17,400	56
	C_{14}	0.00176	32,100	95
	C_{16}		61,700	170

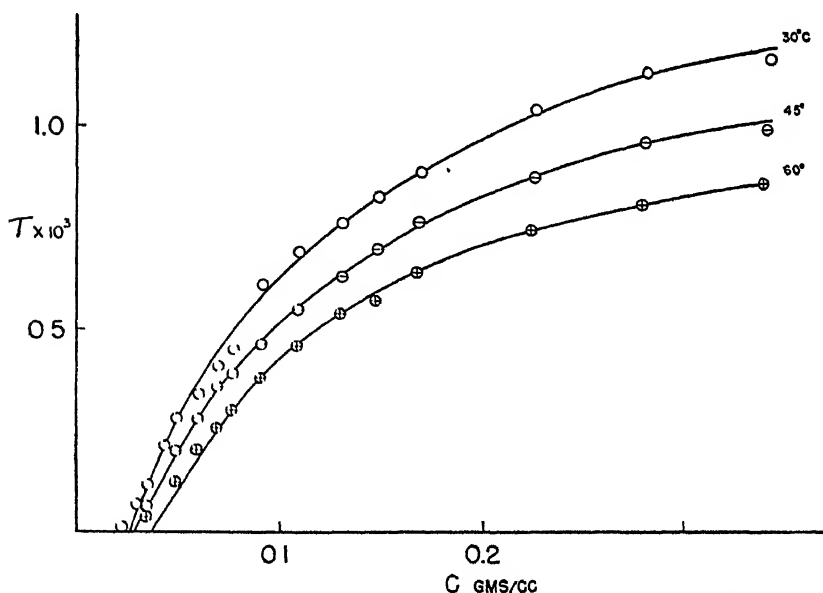


FIGURE 9 Turbidity of *n*-dodecyl trimethylammonium bromide in 0.03403M KBr as a function of temperature.

TBIC from 0.00341 to 0.01476 and N from 75 to 95. Thus, in one instance we increase the TBIC by 14.6 per cent while in the other we increase it by 333 per cent. It is therefore not surprising that N increases only by 5.5 per cent for the *n*-decyl soap and increases 26.7 per cent for the *n*-tetradecyl soap.

In FIGURE 9, are plotted the results of turbidity measurements carried

out with solutions of *n*-dodecyl trimethylammonium bromide in 0.03403M KBr at different temperatures. Molecular weights have been calculated on the assumption that H is independent of temperature. Our differential refractometer, in its present form, cannot be used to determine experimentally at high temperatures the factor $\left(\frac{\mu - \mu_0}{c}\right)^2$ which occurs in H . A calculation employing the Lorenz-Lorentz formula and the density data of Scott and Tartar¹⁷ was carried out to determine the approximate change in this factor with temperature. According to this computation, $\left(\frac{\mu - \mu_0}{c}\right)^2$ decreases by 2.5 per cent as the temperature goes from 30° to 60°C. Varying just the last figure of the density data by one changes the percentage decrease to 2.5 ± 3.3 per cent. This shows that the calculation is very uncertain, consequently we have not taken into account any temperature variation of H .

TABLE 4
n-DODECYL TRIMETHYLAMMONIUM BROMIDE IN 0.03403M KBr

Temperature	Critical concentration	MW
30°C.	0.00848M	20,200
45	0.00931	17,000
60	0.01201	16,400

As seen in TABLE 4, the critical concentration for micelle formation increases with temperature. This is in agreement with the findings of other workers employing different methods of investigation.^{18,19}

It is evident that the effect of temperature changes is not large.

Sketch of a Theory of the Micelle

Although it is not possible, at this time, to present a finished quantitative picture of the micelle, its constitution and the meaning of the constants characterizing its size and its appearance, I believe that the following very incomplete considerations may prove to be of some help, in the meantime.

The monomeric ions we have to deal with in our case consist of a hydrocarbon chain which carries a charge at one end. If we want to make a micelle out of a larger number of such ions, which has the general form proposed by McBain and adopted by Harkins and co-workers as a result of their X-ray work, we see that two different kinds of energy will come into play. We shall gain energy because we remove a number of hydrocarbon tails from the surrounding water and bring them in contact with each other in the micelle. The molecular forces of importance in this process are all short range forces often indicated as van der Waal's forces. At the same time, however, we have to bring the charged ends of the monomer nearer to each

other until they fill both the flat ends of the sandwich-micelle with a surface density ultimately determined by the space requirements of the monomer tail. This process requires energy and now the forces we have to overcome are long range electrical forces. The interplay between short range and long range forces seems to be responsible for the structure of the micelle, as can be seen from the following calculation.

Suppose a circular disk of radius a covered with a constant surface density of electricity σ . The total charge of this sheet will be proportional to σa^2 and the potential at the rim of the disk will be proportional to σa . If we add to this disk an additional ring of thickness da which has the surface $2\pi a da$ and the charge $2\pi \sigma a da$, we will do work against the coulomb forces which is proportional $\sigma^2 a^2 da$. The total electrical work involved in building up the disk is seen to be proportional to a^3 and, since the number of molecules involved in the micelle arrangement is proportional to a^2 , we come to the following conclusion: in order to build a micelle containing N molecules, we have to supply an electrical energy

$$W_e = N^{3/2} w_e \quad (14)$$

in which w_e is a fundamental electrical energy left unspecified for the moment. The main point is that, due to the long range character of the electrical forces, the energy involved increases faster than the number of molecules. We can assume, on the other hand, that the energy gained by bringing N hydrocarbon tails in contact (which involves only short range molecular forces) can be represented as

$$W_m = -N w_m \quad (15)$$

with the introduction of another fundamental molecular energy w_m . If we draw the curve for the total energy $W = W_e + W_m$ as a function of the number of molecules N , it is seen that it has a minimum for a certain value $N = N_0$ and, at this point, the energy W_0 of the micelle is negative. This means that the micelle is more stable than N_0 separate molecules and that work is required in order to either increase or decrease the equilibrium number N_0 . From the equation $dW/dN = 0$ and the expression for W , it follows that

$$N_0^{1/2} = \frac{2}{3} \frac{w_m}{w_e} \quad (16)$$

$$W_0 = -N_0 \frac{w_m}{3}.$$

Accepting these relations, we see that we can determine the 2 fundamental constants w_m and w_e of the micelle, if we know N_0 and W_0 . We can determine N_0 from light scattering. In the case of dodecylamine hydrochloride

in water, N_0 is 55. From the first equation of (16) it follows immediately that

$$\frac{w_m}{w_e} = 11.1. \quad (17)$$

If we accept the reasoning in the preceding paragraph concerning the application of the mass action law, we can also determine W_0 . According to thermodynamical principles, the natural logarithm of the equilibrium constant K multiplied with kT is a measure for the free energy difference between the micelle and the equivalent number of free molecules. Since $K = c_0^{N_0-1}$ we come to the conclusion that

$$W_0 = -N_0 \frac{w_m}{3} = (N_0 - 1)kT \ln c_0 \quad (18)$$

(in which c_0 is of course to be expressed in mol-fractions). If we neglect the difference between N_0 and $N_0 - 1$, we find

$$w_m = -3 kT \ln c_0. \quad (19)$$

In our case, c_0 is 1.3110^{-2} moles per liter, so we arrive at

$$w_m = 25 kT \text{ and } w_e = 2,2 kT. \quad (20)$$

We now can, at once, draw a conclusion about the question whether we shall have a broad distribution of micelle sizes or not.

In the vicinity of $N = N_0$, the micelle energy W can be developed in powers of $N - N_0$. Doing this, we find

$$W = W_0 \left[1 - \frac{3}{4} \left(\frac{N - N_0}{N_0} \right)^2 \right]. \quad (21)$$

The average natural fluctuations in size will involve energy differences kT . Assuming $W - W_0 = kT$ and remembering that $W_0 = -N_0 \frac{w_m}{3}$, it follows from (21) that the average natural fluctuation of N_0 is

$$N - N_0 = \sqrt{4N_0 \frac{kT}{w_m}}.$$

With $N_0 = 55$ and $w_m/kT = 25$ we have

$$N - N_0 = 2.97$$

which means that the micelles containing each in the average 55 molecules, will have a narrow distribution which practically does not cover much more than the range from 52 to 58.

Next, we want to investigate whether the values found in EQUATION 20 for w_m and w_e are such that we can reasonably expect values like that.

First of all, with respect to w_m , we shall guess that it should not be very different from the heat of vaporization of a dodecane molecule from the liquid. According to the measurements of the National Bureau of Standards, in the American Petroleum Institute Research Project 44, the heat of vaporization, at 25°C., of nonane is 11.099 Kcal/mole and the increase for one CH_2 -group is 1.18 Kcal/mole. Extrapolating to dodecane, this corresponds to a heat of vaporization of 14.63 Kcal/mole. Assuming for RT a value of 600 cal., this corresponds to 24.4 kT to be compared with 25 kT . On the other hand, w_e can be calculated for a disk covered with a constant density of electricity in a medium of dielectric constant D . It is found that

$$w_e = \frac{4}{3} \sqrt{\frac{2}{\pi}} \frac{\epsilon^2}{D\sqrt{\omega}} \quad (22)$$

in which ϵ is the electronic charge and ω the surface occupied by one monomer. Taking a value of 27 Å² for ω , the electronic charge for ϵ , and assuming for the effective dielectric constant the average of that of water and that of a hydrocarbon, we obtain $w_e = 2.8 kT$. This is to be compared with 2.2 kT .

We can also apply relation (18) in order to make an estimate of the influence of the number of C-atoms in the chain on the critical concentration. Calling ν the number of C-atoms, we know that w_m can very adequately be represented by a linear function of ν . But, if this is so, EQUATION 18 predicts that c_0 will be proportional to an exponential function of $H\nu$, in which the coefficient of ν is the energy gain per CH_2 -group in going over from free monomer molecules in water to their associated form in the micelle, divided by 3 kT . Dr. Harkins and his co-workers have published a table of the critical concentration of fatty acid soaps covering the range from $\nu = 7$ to $\nu = 14$. The relation is indeed exponential and a good representation of the experimental values is given by the relation

$$\ln c_0 = 4.811 - 0.714\nu \quad (23)$$

(in which c_0 is taken in mols/liter). This means that the energy equivalent of one CH_2 -group is 0.714. 3 $kT = 2.14 kT$ or in molar quantities 2.14 $RT = 1280$ cal. To be confronted with this value is the observation that the molar heat of vaporization of hydrocarbons increases 1180 cal. with every added CH_2 -group.

It remains to be seen what an added electrolyte will do to the micelle. Since, in an electrolyte solution, every charge will be surrounded by an excess of ions of opposite sign, its electrical actions will be screened out for larger distances. In the theory of strong electrolytes, the characteristic distance is the thickness of the ionic layer which, in water and for a uni-univalent electrolyte, is 100 Å for a 0.001 molar and 10 Å for a 0.1 molar solution. We can say, at once, that an added electrolyte will screen the action of the charges on the micelle, will reduce the electrical work W_e , and therefore will increase the equilibrium size of the micelle as has been

observed. The concentration of surface charges on the micelle (one electronic charge on a surface of 27 \AA^2) is so high that the potential in the neighborhood of this surface will certainly exceed 25 millivolts. This value of 25 millivolts corresponds to kT divided by the charge of the electron and is a kind of fundamental unit for the potential. As long as the potentials which the ions encounter are small compared with 25 millivolts, the solution can be characterized by its ionic strength and the variations in number of positive and negative ions from their equilibrium values can be considered as small. As soon, however, as those potentials are a few times 25 millivolts there will be a huge difference between the concentrations of oppositely charged ions. Some considerations led me to a value of 150 to 200 millivolts, 6 to 8 times the fundamental potential for the potential at the surface of the micelle. This would mean that here the positive ions are so strongly repulsed that their concentration is $e^6 = 400$ to $e^8 = 3000$ times smaller than at larger distances. Perhaps this is the way to understand the fact that foreign ions of the same sign as the micellar ion are of very minor importance for its behavior and that the principle of "ionic strength" (which is essentially linked with the existence of potentials small as compared with 25 millivolts) does not hold.

References

1. MCBAIN. 1913. *Trans. Faraday Soc.* 9: 99.
2. MCBAIN. 1944. *Alexander's Colloid Chemistry*. 5. Reinhold. New York.
3. HARTLEY. 1936. *Aqueous Solutions of Paraffin-chain salts*. Hermann et Cie. Paris.
4. HARTLEY, COLLIE, & SOMIS. 1936. *Trans. Faraday Soc.* 32: 795.
5. HARTLEY & RUNNICES. 1938. *Proc. Roy. Soc. London*. A168: 420.
6. MILLER & ANDERSSON. 1942. *J. Biol. Chem.* 144: 475.
7. HAKALA. 1943. *Dissertation*. Univ. of Wisconsin. Quoted in *Ann. N. Y. Acad. Sci.* 1945. 46(5): 326.
8. VETTER. 1947. *J. Phy. & Colloid Chem.* 51: 262.
9. GONICK & MCBAIN. 1947. *J. Am. Chem. Soc.* 69: 334.
10. MCBAIN. 1942. *Advances in Colloid Science*. 1: 124. Interscience Publishers.
11. HARKINS. 1948. *J. Chem. Phys.* 16: 156.
12. DEBYE, P. 1948. *J. App. Phy.* 52: 260.
13. DEBYE, P. 1947. *J. Phys. & Colloid Chem.* 51: 18.
14. EWART, ROE, DEBYE, & MCCARTNEY. 1946. *J. Chem. Phys.* 14: 687.
15. EINSTEIN. 1910. *Ann. Physik*. 33: 1275.
16. CORRIN & HARKINS. 1947. *J. Am. Chem. Sci.* 69: 683.
17. SCOTT & TARTAR. 1943. *J. Am. Chem. Soc.* 65: 692.
18. KLEVEN. 1948. *J. Phys. Colloid Chem.* 52: 130.
19. WRIGHT, ABBOTT, SIVERTZ, & TARTAR. 1939. *J. Am. Chem. Soc.* 61: 549.

THE ROLE OF HYDRATION IN THE DEBYE-HÜCKEL THEORY

BY R. A. ROBINSON AND R. H. STOKES*

Department of Chemistry, Yale University, New Haven, Connecticut and Physical Chemistry Laboratories, Cambridge, England

In 1923, Debye and Hückel derived an expression for the activity coefficient of an electrolyte in which allowance was made for the finite size of the ions, an expression which, in the case of an aqueous solution at 25°, takes the form:

$$-\log f = 0.5092 z_1 z_2 \sqrt{\mu} / (1 + 0.3286 \text{ \AA} \sqrt{\mu}). \quad (1)$$

The quantity \AA , which is usually of the order of 3–5 \AA , is interpreted as a mean ionic diameter, or the closest distance of approach of the ions. Although the most suitable value of \AA does depend on the concentration range over which the experimental data are fitted to EQUATION 1, there is general agreement as to its approximate value; thus, for sodium chloride, values ranging from 3.6 to 4.4 \AA have been obtained.¹ Taking an average of 4.0 \AA , this mean ionic diameter is in marked contrast to the sum of the crystallographic radii, 2.76 \AA . For magnesium chloride, the \AA value is 5 \AA and the sum of the crystallographic radii is only 2.46 \AA . There is, therefore, a large amount of space not accounted for, a fact which suggests very strongly that either or both of the ions are "hydrated" in the sense that an ion and some water molecules act as a unit in so far as the \AA term of EQUATION 1 is concerned.

A second line of evidence for hydration is obtained from the thermodynamic properties of very concentrated solutions. In many of these solutions, the activity coefficient reaches extraordinarily high values, for example, at a concentration of 10M the activity coefficients of calcium chloride, lithium chloride and perchloric acid are 43.0, 9.40 and 30.9, respectively. Several years ago, Scatchard² showed that a solution of sucrose in water behaved as an almost ideal solution when it was treated as one of the penta-hydrated solute in water. If by m and φ we denote the molality and osmotic coefficient of the solution without regard for the hydration, *i.e.*, the values as usually calculated, and m' and φ' the corresponding quantities in which account is taken of hydration, then

$$m' = m / (1 - 0.018 \times 5 m),$$

and from the definition of φ ,

$$-55.51 \ln a_w = m\varphi = m'\varphi',$$

it follows that:

$$\varphi' = \varphi (1 - 0.090 m).$$

* Imperial Chemical Industries Research Fellow, University of Cambridge, Cambridge, England.

The magnitude of the hydration effect may be seen from the values for sucrose, which show that the deviations from ideality can be accounted for satisfactorily on the basis of hydration (TABLE 1).

By introducing the idea of hydration for electrolytes and converting the osmotic and activity coefficients into the corresponding quantities for the hydrated electrolyte, a reduction of φ to a more reasonable magnitude will occur, and it is easy to demonstrate that the amount of hydration which it is necessary to postulate is also of reasonable magnitude.

Returning now to the Debye-Hückel equation, if EQUATION 1 is an expression for the activity coefficient of an assembly containing not only ions but water in some way "hydrated," it is surely not correct to expect this activity coefficient (and the corresponding ionic free energy) to agree, except in very dilute solution, with the activity coefficient obtained from an experiment in which the calculation is performed on the "unhydrated" ions. Rather should we expect better agreement if the Debye-Hückel free energy were equated to the free energy of the ions including their "hydrate" water.

TABLE 1

m	0.1	0.2	0.5	1.0	1.5	2.0
φ	1.008	1.016	1.041	1.088	1.139	1.189
φ'	0.999	0.998	0.994	0.991	0.985	0.975

This problem has already been considered by Bjerrum³ and by Scatchard⁴ and recently by the present authors.⁵ We may denote by primed symbols quantities calculated on the assumption that the kinetic entities produced by one molecule of solute are ions associated with a total of n molecules of "bound" or "hydrate" water, unprimed symbols being used for the corresponding quantities with disregard of "hydration." Moreover, we shall assume that the activity coefficient of the "hydrated" electrolyte is given (on the mol fraction scale) by

$$-\log f' = 0.5092 z_1 z_2 \sqrt{\mu} (1 + 0.3286 \sqrt{\mu}); \quad (2)$$

i.e., we shall not allow any term linear in the concentration to take account of variation in dielectric constant, etc. Then the activity coefficient of the electrolyte, on the molality scale, calculated without regard for "hydration," *i.e.*, the activity coefficient as usually expressed, can be shown to be:⁵

$$\log \gamma = -0.5092 z_1 z_2 \sqrt{\mu} / (1 + 0.3286 \sqrt{\mu})$$

$$-n/\nu \cdot \log a_w - \log [1 - 0.018(n - \nu)m], \quad (3)$$

ν being the number of ions into which one molecule dissociates. Since μ is in volume units, it is not affected by the hydration hypothesis. We have recently⁵ examined the application of this equation and find that it holds

with remarkable accuracy for 36 electrolytes up to concentrations in some cases as high as 5 *m*, using values of *n* ranging from 0.6 to 20. In general, the lower the value of *n*, the higher is the concentration to which the equation can be applied. The *n* value is independent of concentration.

TABLE 2 illustrates the use of the equation in the case of calcium chloride, using values of *n* = 12.0 and \bar{a} = 4.73 to obtain the calculated activity coefficients in EQUATION 3; these are compared with the observed values of McLeod and Gordon⁶ in dilute solution and of Stokes⁷ at higher concentrations.

In the case of the other electrolytes which we have investigated, good agreement is also found. We thus have a two-parameter equation which needs no term linear in the concentration to account for the change in the dielectric constant and yet is equally, if not more, successful in representing

TABLE 2
OBSERVED AND CALCULATED ACTIVITY COEFFICIENTS OF CALCIUM CHLORIDE AT 25°

<i>m</i>	$\gamma_{\text{obs.}}$	$\gamma_{\text{calc.}}$ (EQN. 3)	$\gamma_{\text{calc.}}$ (EQN. 4)
0.0016	0.8640	0.8636	0.8636
.0064	.7667	.7672	.7674
.0144	.6947	.6955	.6957
.0256	.6399	.6408	.6410
.0400	.5974	.5984	.5987
.0576	.5642	.5648	.5651
.0784	.5377	.5377	.5380
.1	.518	.518	.519
.2	.472	.472	.472
.5	.448	.448	.448
1.0	.500	.500	.499
1.4	.587	.586	.586
1.8	.712	.725	.720

the experimental data for 1:1 salts, while for 2:1 salts (for which the Debye-Hückel equation with an added linear term is not adequate above about 0.3 *m*) EQUATION 3 represents a very considerable improvement.

If we are to ascribe the large values of \bar{a} to hydration, it should be possible to calculate \bar{a} in terms of *n*, *i.e.*, to reduce EQUATION 3 to a one-parameter equation. To this end, we have adopted the hypothesis of Bernal and Fowler⁸ that the chloride, bromide, and iodide ions are not hydrated. The hypothesis that cations hydrate more readily than anions has recently obtained some experimental support. In FIGURE 1, we reproduce a graph from a paper by Robinson and Levien⁹ in which are contrasted the activity coefficient curves of salts of 1:2, 2:1, 1:3, 3:1, 1:4, and 4:1 valence type. In each case, if the high charge resides on the anion, a low activity coefficient curve is obtained. If the high charge is on the cation, then the curve is high, characteristic of a highly hydrated salt. This suggests that hydration is mainly a cationic phenomenon and increases with the cation charge.

We shall, therefore, assume that the anion is not hydrated and that the anionic portion of \bar{a} is equal to the crystallographic radius, r_- . The cationic contribution could be obtained if we knew the volume of the hydrated cation. Since the apparent molal volume of the salt is the molal volume which the ions would occupy were there no change in volume of the water in proceeding from pure solvent to the solution, and, as the volume of a water molecule is 30 cu. Å, we should be able to approximate the volume of the hydrated ion as $(30 n + V_+)$ where V_+ is the apparent molal volume of the cation in cu. Å per ion. Fortunately V_+ is usually small compared with 30 n . Furthermore, the apparent molal volumes of the chlorides, bromides, and iodides of rubidium and cesium, salts which, owing to the large size of both the cation and the anion, are either not hydrated or else hydrated to a very small extent, are found to be given at about 1M by: $V_{app} = 6.47(r_+^3 + r_-^3)$ in cu. Å per ion, r_+ and r_- being the crystallographic radii. We therefore

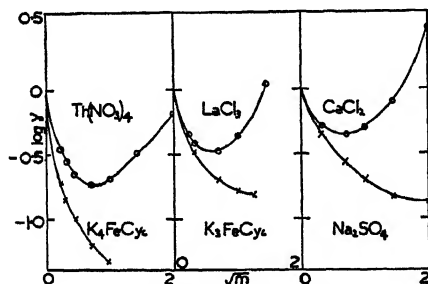


FIGURE 1. Activity coefficients of salts of various types.

assume that the anionic apparent molal volume is $6.47 r_-^3$ for all halides and that of the cation of a halide salt is $V_+ = (V_{app} - 6.47 z_1 r_-^3)$, a quantity which can be calculated from density data. In this expression, z_1 is the cation valence. We thus obtain $[3(30 n + V_{app} - 6.47 z_1 r_-^3)/4\pi]^{1/3}$ as the cationic contribution to \bar{a} . The distance of closest approach should therefore be

$$\bar{a} = [3(30 n + V_+)/4\pi]^{1/3} + r_-.$$

Comparison with the values of \bar{a} obtained from EQUATIONS 1 and 2 shows that the calculated value is uniformly higher by approximately 0.7 in the case of a 1:1 salt and 1.3 for a 2:1 salt. This distance may well represent a penetration of the anion into the hydration layer.¹⁰ Taking this into account, we obtain the one-parameter equation

$$\log \gamma = -0.5092 z_1 z_2 \sqrt{\mu} / \{1 + 0.3286 \sqrt{\mu} [3(30 n + V_+)/4\pi]^{1/3} + r_- - \Delta\} - n/\nu \cdot \log a_w - \log [1 - 0.018(n - \nu)m] \quad (4)$$

for an electrolyte in water at 25°, where $\Delta = 0.7$ for a 1:1 electrolyte and 1.3 for a 1:2 electrolyte.

We have applied this equation to the data for the chlorides, bromides, and iodides of hydrogen, lithium, sodium, potassium, rubidium, magnesium, calcium, strontium, and barium and for manganese, ferrous, cobalt, and nickel chloride. The success with which this equation can be used may be seen from TABLE 3.¹

TABLE 3

<i>Electrolyte</i>	<i>n</i>	<i>Range up to</i>	<i>Average difference in γ</i>	<i>Maximum difference in γ</i>
HCl	7.3	1.5m	0.002	0.003
HBr	8.6	1.0	.001	.002
HI	10.6	0.7	.0015	.002
LiCl	6.5	2.0	.0025	.005
LiBr	7.1	2.5	.004	.006
LiI	10.0	0.5	.002	.003
NaCl	3.5	5.0	.0035	.004
NaBr	4.15	4.0	.002	.005
NaI	5.05	3.0	.002	.004
KCl	1.9	4.0	.002	.003
KBr	2.05	4.0	.002	.003
KI	2.45	4.0	.002	.003
RbCl	1.25	2.0	.002	.004
RbBr	0.9	2.0	.0015	.005
RbI	0.6	2.0	.0035	.007
MgCl ₂	13.9	1.4	.001	.005
MgBr ₂	17.0	1.0	.002	.005
MgI ₂	20.0	0.7	.004	.005
CaCl ₂	11.9	1.4	.0005	.001
CaBr ₂	14.0	1.0	.004	.006
CaI ₂	17.0	1.0	.002	.007
SrCl ₂	10.8	1.8	.001	.005
SrBr ₂	12.4	1.4	.002	.005
SrI ₂	15.5	1.0	.0015	.003
BaCl ₂	8.4	1.8	.004	.009
BaBr ₂	10.3	1.8	.002	.006
BaI ₂	14.7	1.0	.0035	.006
MnCl ₂	11.4	1.0	.002	.004
FeCl ₂	12.1	1.4	.001	.002
CoCl ₂	13.0	1.0	.001	.002
NiCl ₂	13.1	1.4	.001	.003

IN TABLE 2, we have tabulated values for calcium chloride calculated by EQUATION 4 with $n = 11.9$ and $\bar{a} = 4.75$. In this case, EQUATIONS 3 and 4 apply equally well but, in general, the two-parameter equation is somewhat superior in representing the experimental data.

It is not easy to test EQUATION 4 by application to salts of higher valence type, but we have investigated the data for lanthanum chloride, for which salt we have obtained activity coefficients from 0.1 to 2 *m*. Unfortunately, we do not know with certainty the value at 0.1 *m*, but if we arbitrarily assign the not unreasonable value of 0.314 to this concentration, we find that good agreement with the experimental data is obtained if in EQUATION 4 we put

$n = 18.2$ (in which case $\text{\AA} = 4.90 \text{\AA}$) and $\Delta = 1.7$, as is shown in the following comparison:

m	$\gamma(\text{obs})$	$\gamma(\text{calc})$
0.1	0.314	0.314
.2	.275	.276
.3	.264	.265
.5	.268	.267
.7	.288	.287
1.0	.344	.345

Effect of Temperature. To test out EQUATION 4 still further, we have applied it to hydrochloric acid,¹¹ hydrobromic acid,¹² sodium chloride,¹³ sodium bromide,¹⁴ potassium chloride,¹⁵ and potassium bromide,¹⁶ electro-

TABLE 4

Temp.	HCl			HBr			NaCl			NaBr			KCl			KBr		
	n	Av.	Max.	n	Av.	Max.	n	Av.	Max.	n	Av.	Max.	n	Av.	Max.	n	Av.	Max.
0°	7.70	3	7	8.90	2	4	2.85	5	11	3.30	7	22	1.30	4	14			
10	7.56	3	6	8.80	2	3	3.20	5		83.75	5	181	1.65	2	5			
20	7.35	3	5	8.63	2	3	3.43	3		54.00	4	51.88	2	3				
25	7.30	3	9	8.60	1	2	3.50	2		34.15	2	51.90	2	3		2.05	2	3
30	7.15	2	5	8.45	2	4	3.60	1		34.22	3	132.00	2	5				
40	6.90	2	4	8.30	1	2	3.70	1		24.28	4	172.08	2	3				
50	6.70	2	3	8.15	1	2												
60	6.50	2	4	7.85	5	6	3.90	2	5							2.60	3	5
70							3.87	3	4							2.70	3	8
80							3.85	4	7							2.75	3	8
90							3.80	2	4							2.80	4	11
100							3.77	2	5							2.82	3	10
Range	0.1-2 m			0.1-1 m			0.1-3 m			0.1-3 m			0.1-4 m			0.1-5 m		

lytes whose activity coefficients have been measured over a temperature range.

We have thrown all the temperature effect on the variation of n and the numerical coefficients of the Debye-Hückel term with temperature, although a more accurate but very laborious computation would take account of the effect of temperature on the volume of the water molecule, the apparent molal volume of the cation, and "penetration distance." TABLE 4 gives our results, the columns headed Av. and Max. giving the average and maximum differences between the third decimal places of the observed and calculated activities coefficients.

The variation of n with temperature is shown in FIGURE 2. It is tempting to speculate that there may be a limiting value of $n = 4$ at high temperature, but the data available are so meager that further comment would not be justified. We may, however, call attention to the large amount of work still to be done on the activity coefficients of simple salts at different tem-

peratures. In spite of the approximate nature of our calculations, EQUATION 4 seems to be applicable over a wide temperature range.

Partial Molal Heat Content. If EQUATION 4 can account for activity coefficients over a temperature range, it should also account for partial molal heat contents. By differentiating EQUATION 4 with respect to the temperature and substituting the proper numerical values for various parameters, we have obtained the equation:

$$\begin{aligned} \bar{L}_2 = [774 \sqrt{c} + 169.6 \text{ } \text{\AA} c + 4.141 \times 10^5 d\sqrt{c}/dT] / (1 + 0.3286 \text{ } \text{\AA} \sqrt{c})^2 \\ - n\bar{L}_1 + \partial n / \partial T \cdot [4.066 \times 10^5 \log a_w - 6.358 \times 10^3 m / \{1 - 0.018(n - \nu)m\} \\ - 3.249 \times 10^5 c / k^{2/3} (1 + 0.3286 \text{ } \text{\AA} \sqrt{c})^2] \end{aligned} \quad (5)$$

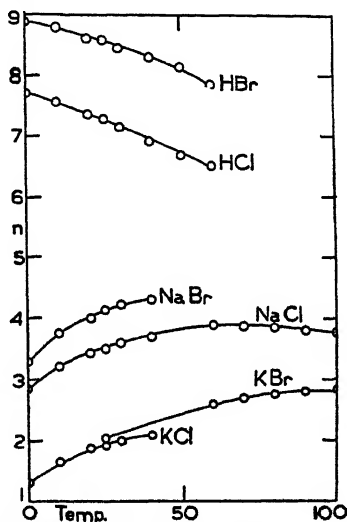


FIGURE 2. Effect of temperature on the hydration number.

where $k = 3(30n + V_+)/4\pi$, valid for a 1:1 electrolyte in water at 25°. We have tested this equation by calculating \bar{L}_2 for hydrochloric acid and sodium chloride, assuming that $\partial n / \partial T = -0.0200$ and $+0.0167$ respectively. The calculated values are compared in TABLE 5 with those measured by Sturtevant¹⁷ and by Robinson,¹⁸ and a graphical comparison is made in FIGURE 3. The agreement is good. It might perhaps be better if we were to engage in a more detailed computation but we are content to show that two parameters, n and $\partial n / \partial T$, suffice to give a fair representation of the behavior of the heat content function.

Mixed Electrolyte Solutions. For a solution of two electrolytes at constant total ionic strength, Harned's rule¹⁹ states that the logarithm of the activity coefficient of one electrolyte is a linear function of the molality of that electrolyte, i.e.,

$$\log \gamma_1 = \log \gamma_{1(0)} - \alpha_{12}m_2 = \log \gamma_{0(1)} + \alpha_{12}m_1$$

TABLE 5

m	\bar{L}_2 (NaCl), cal./mole		\bar{L}_2 (HCl), cal./mole	
	obs.	calc.	obs.	calc.
0.1	102	95	202	176
.2	90	87	273	250
.5	-10	34	430	392
1.0	-188	-87	645	607
1.5	-343	-203	853	819
2.0	-466	-347	1055	1040
2.5	-556	-490	1269	1341
3.0	-626	-629	1484	1631
3.5	-671	-774	—	—
4.0	-688	-920	—	—
$\partial n/\partial T$	+0.0167		-0.0200	

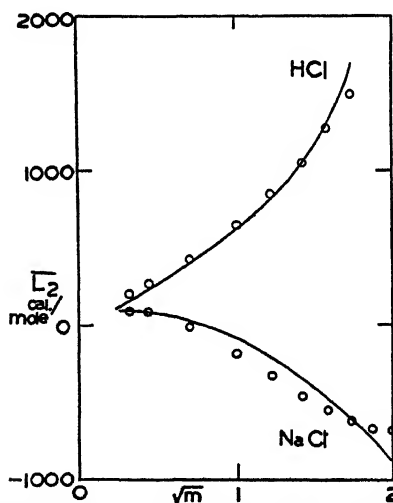


FIGURE 3. Partial molal heat contents at 25°. Curves show calculated, points experimental values.

where $\gamma_{1(0)}$ is the activity coefficient of species (1) in the absence of species (2) and $\gamma_{0(1)}$ refers to species (1) in a solution of species (2) only. By an extension of the ideas already developed, we can derive the equation

$$\log \gamma_1 = \log f'_1 - n_1/2 \cdot \log a_w - \log [1 - 0.018(n_1 m_1 + n_2 m_2 - 2m)] \quad (6)$$

for the special case of a mixture of 1:1 electrolytes. n_1 and n_2 are the hydration numbers and m_1 , m_2 the molalities of species (1) and (2) respectively, m being the total molality.

When $m_1 = m$, $m_2 = 0$:

$$\log \gamma_{1(0)} = \log f'_{1(0)} - n_1/2 \cdot \log a_{w(1)} - \log [1 - 0.018(n_1 - 2)m]$$

and when $m_1 = 0$, $m_2 = m$

$$\log \gamma_{0(1)} = \log f'_{0(1)} - n_1/2 \cdot \log a_{w(2)} - \log [1 - 0.018(n_2 - 2)m].$$

But $\log \gamma_{0(1)} = \log \gamma_{1(0)} - \alpha_{12}m$,

$$\therefore \alpha_{12}m = \log f'_{1(0)}/f'_{0(1)} - n_1/2 \cdot \log a_{w(1)}/a_{w(2)} - \log [1 - 0.018(n_1 - 2)m]/[1 - 0.018(n_2 - 2)m].$$

n_1 and n_2 are known from measurements on the individual electrolytes; $a_{w(1)}$ and $a_{w(2)}$ are the water activities of solution of species (1) and (2) respectively, each at a concentration, m . Moreover, we have for an aqueous solution at 25°:

$$\log f'_{1(0)}/f'_{0(1)} = 0.5092[\sqrt{c_2}/(1 + 0.3286\sqrt{c_2}) - \sqrt{c_1}/(1 + 0.3286\sqrt{c_1})]$$

where $f'_{1(0)}$ and $f'_{0(1)}$ are activity coefficients (on the mol fraction scale) of the hydrated electrolyte (1) in a solution containing species (1) only and species

TABLE 6

m	<i>LiCl</i>		<i>NaCl</i>		<i>KCl</i>	
	$\alpha_{12}(\text{calc.})$	$\alpha_{12}(\text{obs.})$	$\alpha_{12}(\text{calc.})$	$\alpha_{12}(\text{obs.})$	$\alpha_{12}(\text{calc.})$	$\alpha_{12}(\text{obs.})$
0.1	0.006	0.001	0.030	0.043	0.043	0.077
.5	.006	.006	.033	.037	.047	.062
1.0	.008	.005	.037	.032	.052	.056
1.5	.009	.005	.041	.032	.055	.055
2	.010	.005	.045	.031	.062	.057
3	.011	.004	.053	.031	.073	.062
4	.008	-.003	.055	.030	.075	.066

(2) only, respectively, and c_1 and c_2 are concentrations on the volume scale. This term is usually negligible. In any case, it can be calculated from density data. It should be noted, however, that we have not allowed for any change in λ . Disregarding this for the time, α_{12} can now be calculated from n_1 , n_2 , $a_{w(1)}$ and $a_{w(2)}$. This has been done for (1) = HCl and (2) = LiCl, NaCl, and KCl with the results shown in TABLE 6, the observed values being taken from the tables of Harned and Owen.¹⁹ The agreement is not all that we could wish. Perhaps we have not placed sufficient emphasis on the difficult question of what value of λ should be used in these mixed electrolyte solutions. Moreover, we may be incorrect in taking over for mixed solutions the values of n_1 and n_2 used for the simple solutions, but we may be content that we can predict at least the order of α_{12} . In particular, we can say that α_{12} will be greater the greater the difference in n_1 and n_2 , the hydration numbers of the two electrolytes. Even as an approximation, EQUATION 6 will have its uses. Thus, it predicts that at zero acid concentration and 1 m lithium, sodium and potassium chloride, $\gamma_{\text{HCl}} = 0.794, 0.743,$

and 0.716 respectively, compared with 0.800, 0.751, and 0.711; and, at 0.1M, 0.795, 0.791, and 0.789, compared with 0.796, 0.789, and 0.782.

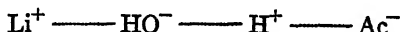
The most serious drawback to EQUATION 4 is the lack of additivity of the n values. Thus, $n_{\text{LiBr}} - n_{\text{LiCl}} = 0.6$ and $n_{\text{NaBr}} - n_{\text{NaCl}} = 0.65$, whilst $n_{\text{KBr}} - n_{\text{KCl}} = 0.15$. We have no adequate explanation to offer for this curious behavior, although it might well be suggested that the omission, from EQUATION 2, of a term linear in the concentration to allow for the change in dielectric constant may be too drastic a procedure and that such a term may be necessary. If this were done, the coefficient of the linear term would have to be smaller than the one used in the extended Debye-Hückel equation and the value of n would have to be less than the one used in EQUATION 4, because both effects raise the activity coefficient; EQUATION 4 is not incompatible with such a term linear in the concentration. We have not pursued this line because it leads to great difficulty in computation and we have been more interested in finding how far EQUATION 4 in its simplest form can be used.

Properties of Very Concentrated Solutions. EQUATION 4 holds up to concentrations as high as 5m for some salts; indeed, it is valid to higher concentrations than we might expect. Moreover, when the concentration is too high for the equation to hold, we notice that in nearly all cases the equation predicts values of the activity coefficient higher than those observed. As the amount of the water relative to the solute decreases, it is to be expected that the degree of hydration will diminish, that is, we should use a decreasing value of n in EQUATION 4 and, therefore, the high calculated values of the activity coefficients are not surprising.

At very high concentrations (about 11m for 1:1 and 7m for 2:1 electrolytes), we have found that the vapor pressure data conform to a Brunauer-Emmett-Teller²⁰ adsorption isotherm and, in particular, to a modification of this isotherm proposed recently by Anderson.²¹ Thus, we are led to visualize the process of solution of a solid salt as the building-up, around the salt ions, of layers of water molecules, the inner layers being the more densely populated and the salt ions retaining some remnants of their crystalline form in a distorted lattice. As the water layers build up, this lattice becomes more and more distorted or the ions with their water molecules function more as separate entities. Ultimately, when sufficient water is adsorbed, we pass into a region of concentration subject to the treatment given above, in which a portion only of the water is considered as "hydrate water."

The Hypothesis of Localized Hydrolysis. Among the salts of alkali metals are to be found a few for which the order of the activity coefficient curves is $\text{Li} < \text{Na} < \text{K} < \text{Rb} < \text{Cs}$, which is the converse of the order found for the chlorides, bromides, nitrates, etc. This class includes the formates, acetates, and hydroxides, the latter being considered as salts of the very weak acid, water. To explain this reversal of order of the activity coefficient

curves, Robinson and Harned²² postulated that the intense fields around a cation of small radius created an induced polarity of the water in the immediate vicinity. In the absence of any proton acceptor, no further effect would be noted. Thus, in the case of lithium chloride where, although the lithium ion is small enough to induce high polarization in a water molecule, nevertheless the chloride ion is a very weak proton acceptor and the activity coefficient curve is "normal." The situation is very different with lithium acetate, where the acetate ion is a moderately good proton acceptor and we may imagine a water molecule forming a weak link between the two ions:



thereby reducing the total ionic strength and decreasing the activity coefficient calculated in the usual way on the assumption of complete dissociation. The effect should depend on two factors: (a) The size of the cation, the effect being more pronounced the smaller the cation. This accords with the experimental observation that the order of the activity coefficient curves is $\text{Li} < \text{Na} < \text{K} < \text{Rb} < \text{Cs}$. (b) The strength of the acid from which the salt is derived, the effect being larger the weaker the acid. The available evidence, meager as it is, supports this conclusion. Thus, whilst γ_{NaX} is greater than γ_{KX} for salts of strong acids such as hydrochloric, hydrobromic, and hydriodic acid, the activity coefficient curves of sodium formate and acetate are slightly below those of the corresponding potassium salts, but sodium hydroxide has an activity coefficient curve markedly lower than that of potassium hydroxide. Data for more salts are required before further progress can be made.

If the idea we have developed is correct, cesium acetate and hydroxide should be normal, since the cesium ion is too large to be hydrated. The very high activity coefficients of cesium hydroxide would then correspond to a high λ value, which in turn would mean that the hydroxyl ion is hydrated. This is consistent with the apparent molal volume of the hydroxides, as was pointed out by Bernal and Fowler.⁸ Furthermore, we may have to admit hydration of the acetate ion.

Summary

It is shown that the departures from the limiting Debye-Hückel equation for the activity coefficient of a salt can be explained as a hydration effect, and a one-parameter equation is proposed in which the effective mean diameter is a function of the hydration number. This equation requires no terms involving the first or higher powers of the concentration and represents the activity coefficient data of 32 electrolytes at 25° within the probable experimental error. It represents the data for six electrolytes over a temperature range and, with the addition of a second parameter, the temperature coefficient of the hydration number, gives a fair representation of the variation

of the partial molal heat content with concentration. It also affords a partial explanation of the behavior of the activity coefficients of mixed electrolyte solutions.

Bibliography

1. BROWN, A. S., & D. A. MACINNES. 1935. J. Am. Chem. Soc. 57: 1356.
2. HARNED, H. S., & L. F. NIMS. 1932. *Ibid.* 54: 423.
3. SCATCHARD, G. 1921. J. Am. Chem. Soc. 43: 2406.
4. BJERRUM, N. 1919. Meddel. Vetensk. Nobelinst. 5:1.
5. 1920. Z. anorg. Chem., 109: 275.
6. SCATCHARD, G. 1925. J. Am. Chem. Soc. 47: 2098.
7. STOKES, R. H., & R. A. ROBINSON. 1948. J. Am. Chem. Soc. 70: 1874.
8. MCLEOD, H. G., & A. R. GORDON. 1946. J. Am. Chem. Soc. 68: 58.
9. STOKES, R. H. 1948. Trans. Faraday Soc. 44: 295.
10. BERNAL, J. D., & R. H. FOWLER. 1933. J. Chem. Phys. 1: 515.
11. ROBINSON, R. A., & B. J. LEVIEN. 1947. Trans. Roy. Soc., N.Z. 76: 295.
12. FRANK, H. S. 1941. J. Am. Chem. Soc. 63: 1789.
13. HARNED, H. S., & R. W. EHLERS. 1933. J. Am. Chem. Soc. 55: 2179.
14. HARNED, H. S., A. S. KESTON, & J. G. DONELSON. 1936. J. Am. Chem. Soc. 58: 989.
15. HARNED, H. S., & L. F. NIMS. 1932. J. Am. Chem. Soc. 54: 423.
16. SMITH, R. P., & D. S. HIRTLE. 1939. J. Am. Chem. Soc. 61: 1123.
17. HARNED, H. S., & C. C. CRAWFORD. 1937. J. Am. Chem. Soc. 59: 1903.
18. HARNED, H. S., & M. A. COOK. 1937. J. Am. Chem. Soc. 59: 1290.
19. JOHNSON, G. C., & R. P. SMITH. 1941. J. Am. Chem. Soc. 63: 1351.
20. STURTEVANT, J. M. 1940. J. Am. Chem. Soc. 62: 58.
21. ROBINSON, A. L. 1932. J. Am. Chem. Soc. 54: 1311.
22. HARNED, H. S., & B. B. OWEN. 1943. *Physical Chemistry of Electrolytic Solutions*. XIV. Reinhold. New York.
23. BRUNAUER, S., P. H. EMMETT, & E. TELLER. 1938. J. Am. Chem. Soc. 60: 309.
24. ANDERSON, R. B. 1946. J. Am. Chem. Soc. 68: 686.
25. ROBINSON, R. A., & H. S. HARNED. 1941. Chem. Rev. 28: 419.

SOME NEW PROCEDURES IN THERMODYNAMIC THEORY INSPIRED BY THE RECENT WORK OF J. N. BRØNSTED

BY VICTOR K. LAMER, OLAV FOSS,¹ AND HOWARD REISS
Department of Chemistry, Columbia University, New York, N. Y.

Introduction

During the past decade, Professor J. N. Brønsted has been engaged in developing a system for presenting in a simple fashion the principles and concepts of thermodynamics. Cast into this form he called the subject "Energetics." Instead of the traditional first and second laws, he proposes two new principles, namely, the "work principle," which is restricted to, and is sufficient for, an exhaustive treatment of all reversible processes, and the "heat and equivalence" principle, which applies similarly to irreversible processes. These principles are expressed in compact analytical form in one equation (4.22) in the following text.

They are introduced as general postulates based upon experience, just as the first and second laws of thermodynamics are introduced.

With the aid of the "work principle," Brønsted achieves in a simple and elegant manner a uniform treatment of all reversible processes on the basis of the concepts of quantity, potential, and "work,"[†] without introducing the concept of heat, or of internal energy. More especially, all reversible thermal processes may be completely described in terms of temperature and entropy. It is only when proceeding to irreversible processes that phenomena occur which require a concept of "heat," which embodies some but not all of the characteristics of the heat concept employed in the two classical laws. Internal energy is recognized and defined. Likewise, the various forms of the characteristic thermodynamic potentials introduced by Gibbs are defined as convenient functions for special applications, but their introduction like that of internal energy is not a necessary part of his system.

The basic ideas of energetics appeared in two monographs^{2,3} in Danish in 1937 and 1939. A brief summary, translated by R. P. Bell, appeared in English in 1940,⁴ and also a short paper⁵ in which the "work principle" was used to derive the equilibrium equations for heterogeneous and homogeneous systems. An important monograph⁷ clarifying criticisms⁶ based upon misunderstanding of this paper⁵ was published in 1941 (in English) under the title, *The Concept of Heat*.

The fundamentals of energetics were incorporated in the second Danish

^{*} Fellow, 1947-8, The Royal Norwegian Council for Scientific and Industrial Research. Present address: Universitetets Kjemiske Institutt, Blindern, Oslo, Norway

[†] Quotes are used to designate Brønsted's use of terms which often differ from the classical meaning of the term.

edition of the textbook of physical chemistry.⁸ In his last monograph⁹ (in Danish), Brønsted introduces the concept of "transport complexes" and applies it together with the "work principle" to the treatment of the reversible aspects (assumed isolable) of steady state processes in thermoelectric, electrochemical, and thermal transpiration cells.

Brønsted's Energetics

The following is a brief perspective view of some of the ideas formulated by Brønsted. Energetics is motivated by the symmetry existing between all of the various extensive energy factors (the quantities) and, likewise, between their conjugate intensive factors (the potentials). For example, in equations like

$$S dT - V dp + \sum n_i d\mu_i + E d\psi + \dots = 0$$

we find entropy and temperature, volume and pressure, number of moles and chemical potential, electric charge and potential etc., always in a conjugate relationship. All of the extensive properties are defined such that they are conserved, except entropy in the case of irreversible processes.

Next it is asserted that the extensive factors, such as volume, mass, electrical charge, etc., and also entropy* always tend to occupy states of lowest accessible potential. Natural (spontaneous) and unnatural (imaginary) processes consist of the movement of the extensive factors between states of different potentials. Each individual transport is a basic process.

Such processes involve, respectively, a positive or a negative *loss of ability to perform useful work*. This loss, in a process where an amount δK_i of an extensive property of the i^{th} kind undergoes a transport from one potential to another, is defined as:

$$\delta A_i \equiv (P_{(1)} - P_{(2)})\delta K_i. \quad (1)$$

Here, $P_{(1)}$ and $P_{(2)}$ are the potentials, conjugate to the quantity K_i ; of the initial and final state of transport, respectively.

When a natural and an unnatural active process are *coupled* and balanced within an infinitesimal difference in net potential, a reversible process results. In the special case when the potentials P_1 and P_2 of a simple process are equal, that process then becomes an equipotential or neutral process, which is obviously a limiting case making contact between the natural and the unnatural classes.

Energetics stresses the necessary presence of a potential difference for a process to occur spontaneously, and makes fundamental use of the existence of coupling between these individual basic processes. It thus focuses more attention upon the physically tangible structures of thermodynamic phenomena than upon the mathematical formulations.

* See the treatment of the heat engine, p. 620.

The "work principle" states that, in reversible processes, the sum A , of the individual A , terms: (*i.e.*, the sum of the "losses of work" in all of the individual processes) is zero:

$$\delta A \equiv \Sigma \delta A, \equiv \Sigma (P_{(1)} - P_{(2)}) \delta K, = 0. \quad (2)$$

In irreversible processes, the sum δA is not zero but positive; there is a positive loss of potential work.

The "equivalence principle" states that:

$$\delta A = T \delta S'', \quad (3)$$

where $\delta S''$ is the amount of entropy produced in the process. $T \delta S''$ is, consequently, the non-compensated heat of Clausius.

We refer the reader to the original publications for the logical operational definitions of such topics as the absolute temperature scale, entropy, and heat in the system of energetics.

In order to distinguish* reversibly transported (conserved) entropy from irreversibly produced entropy, as well as to distinguish reversibly absorbed heat from the non-compensated (irreversibly evolved) heat, Brønsted uses, when necessary, single and double primes, respectively. Thus:

$$\text{Heat reversibly communicated} \quad \delta Q' = T \delta S' \quad (4)$$

$$\text{Heat irreversibly evolved} \quad \delta Q'' = T \delta S''. \quad (5)$$

This notation is adopted in the following text.

Brønsted recognized the necessity of broadening the concept and definition of work beyond the limits it enjoys in the classical presentation in order to achieve the uniform and systematic treatment which he desired for all forms of energy.

In the classical presentation, the element of work DW is an inexact differential defined as

$$DW = P dK. \quad (6)$$

Here, d and D are symbols for exact and inexact differentials; p and K are the conjugate potential and quantity factors; thus $DW = p dV$ for volume work.

On the other hand, Brønsted's "loss of (potential) work," δA , always involves the difference between two potentials. In the reversible case, δA , becomes a function of state and represents the maximum work the i^{th} natural process can perform upon the unnatural process with which it is coupled. In spite of this similarity, δA , should not be confused with the Helmholtz free energy function bearing the same symbol.

For the transport of a finite amount of quantity between states the poten-

* The significance of including terms for the irreversibly produced entropy or the non-compensated heat of Clausius—which Brønsted does in his "equivalence principle"—was pointed out in 1936.¹

It has been stressed recently by TOLMAN & FINE¹⁰ See also ECKART¹¹ and BRIDGMAN.¹² See also DE DONDER & VAN RYSELBERGHE,¹³ and LEAF.¹⁵

tials of which differ infinitesimally, the "loss of work" assumes the form $DA = K dP$ but is not in general integrable. Thus, in the special case, where one of the potentials of a component process can be set equal to zero and the amount of quantity transported is infinitesimal, the "loss of work" assumes the same form and numerical value as the classical EQUATION 6 but should not be confused with it.

Unfortunately, the very use of the terms "work" and "heat" in senses which often differ from the time-honored and specific meanings of the classical presentation leads to confusion no matter how carefully they are defined. Also, it is not easy to look with favor upon a summary replacement of the well-established first and second laws by two new postulates. Only when the advantages of the replacement become evident can one expect general approval.

When one examines the *direct* experimental evidence supporting the postulates of energetics, one will find that it is not abundant, because the attention of investigators has been directed over the past century to the justification of the two laws in their classical form. Although we believe no one who will follow through the logical reasoning of energetics will question the validity of the postulates on these grounds, nevertheless the critical reader and the student approaching the subject for the first time are justified in expecting to be led to energetics from an abundance of direct experimental evidence with which they are familiar.

Consequently, many readers and particularly those who have had access only to the abbreviated presentations available in English have been unsympathetic. Some may have ceased reading before they had had an opportunity to assess the functional value of the ideas embodied in the new principles. As a result, many real contributions contained within the manifold of energetics have been ignored.

One of the objectives of this paper, therefore, is to show that the new postulates are completely equivalent to the classical laws, but that they have in addition certain valuable simplifying didactic merits. One of these is that the new system focuses attention upon the physical structure of the concepts and operations rather than upon the mathematical transformations.

Also, energetics goes beyond thermodynamics in furnishing a generalized model concerning which a universal statement—namely, our *rule of potentials* (see below)—can be enunciated. This does not imply that more information is obtained, but only that a concise and unequivocal form of statement results. Our original plans were to describe Brønsted's system in the manner in which he arranged it. Discussions with colleagues, as our manuscript took shape, demonstrated that confusion resulted from a new terminology. This, and, in addition, the existence of a natural reluctance to base conclusions upon new postulates until their advantages are clearly evident, have led us to reverse the procedure.

Thus, in this paper, we shall abstract and emphasize only those facts which represent tangible contributions by fitting them into the established framework of classical thermodynamics to be used as additional tools.* When the reader becomes satisfied that the new treatment is fully equivalent to the old, and, through use, becomes more confident in its power and simplicity, we hope he will be less reluctant to follow Brønsted's procedure and base all of the reasoning on the new postulates, thus achieving a further gain in didactic simplicity.

To this end, in what follows, we shall avoid objectionable terminology, but have retained the advantageous features of the "spirit" of energetics. They are listed in the section on assessment. In these respects, much of what follows cannot be imputed to Brønsted alone.

Concepts and Definitions

A thermodynamic system is defined as a geometric region whose boundaries may be fixed or variable, and which may contain matter, or energy, or both. The suitable description of such a system depends, in part, upon the specification of the amounts of certain components known as the extensive energy factors, which we shall call *quantities*, following Brønsted. Thus, it is customary to say that the system possesses certain amounts of volume, surface, matter, electric charge, entropy, moles of chemical components, *etc.*

Consider an isolated system, *i.e.*, one which cannot receive quantity from, or lose it to, the regions beyond its geometric boundaries. In addition, we shall at first be concerned only with a system in which no chemical reaction is occurring. Any infinitesimal variation which takes place within this system is limited, either to the *redistribution* of quantity among its physically distinct parts, or to the *production*† of quantity *within* the system, or both. Call those variations, associated with redistribution, *transport* processes.

The comprehensive description of a thermodynamic system requires the numerical specification of another set of entities, known as *intensities* or *potentials*. In this set, we include the familiar parameters, pressure, surface tension, gravitational potential, electrical potential, temperature, and chemical potential *etc.*, which the reader will recognize are each conjugate respectively to the quantities above. During any infinitesimal change, involving the production and redistribution of quantity the potentials remain, sensibly, constant.

The transport of matter, charge, entropy, and moles of chemical constituent, *etc.*, between parts of the system requires no comment. The situation in respect to volume and surface is much the same. However, it is worth while to indicate, clearly, how these latter transports occur.

* For discussions of the relationship between Brønsted's Energetics and traditional Thermodynamics see ROSENBERG, T. H. 1943. *Iysk tidskrift*, 41:1; and HOLTAN JR., H. 1948. *SÆRTRYKK AV TIDSSKRIFT for Kemi, Bergvesen og Metallurgi*, 8: 124-129.

† The only quantity which can be produced, *z.s.*, which is not conserved, is entropy, in irreversible processes. In the reversible case, entropy and thermal processes can be treated in full conformity with other quantities and processes as emphasized by Brønsted.

Imagine a box (FIGURE 1) equipped with a movable partition (cross-hatched) which separates two gases at the pressures p_2 and p_1 , respectively; $p_2 > p_1$. The partition moves to the right, as indicated by the arrow, and the volume, δV , originally on the right of the partition, appears on the left. In this sense, the volume is transported from the region of lower pressure, p_1 , to that of higher pressure, p_2 ; *i.e.*, the potential conjugate to volume is negative pressure. Similarly, we can consider the transport of surface. Consider two films (FIGURE 2) having the tensions, γ_2 and γ_1 ; $\gamma_2 > \gamma_1$; which are distended between two fixed wires (extremes); and a wire free to move as the films demand (center). The center wire will move spontaneously to the

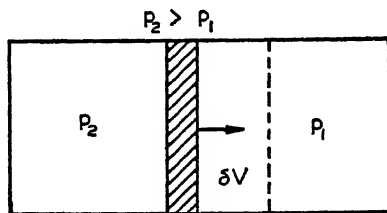


FIGURE 1.

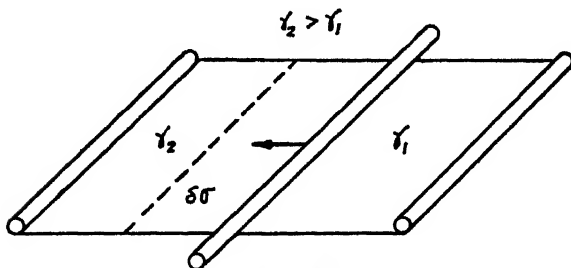


FIGURE 2.

left, as the arrow indicates, and the surface $\delta\sigma$ will be transported from left to right, from the region of higher surface tension to that of lower.

*The Brønsted Principles Derived from the First and Second Laws; the Non-Compensated Heat of Clausius**

Choose an isolated system which is not the seat of chemical reactions, and divide it into localities, in each of which the *potentials* are *uniform*. By combining the first and second laws of thermodynamics, we follow Gibbs and write for the variation δE_j of the internal energy of the j^{th} locality, during any infinitesimal change:

$$\delta E = T\delta S - p\delta V + \gamma\delta\sigma + \sum_i \mu_i \delta n_i + \phi\delta m + \psi\delta e. \quad (4.1)$$

* Since the non-compensated heat measures the amount of energy that became unavailable as useful work in an irreversible process, and is zero for reversible processes, we prefer, for the present, to emphasize this concept, also, for reversible processes, rather than the Brønsted concept "loss of work" (which is defined differently from the traditional "work" of thermodynamics). The non-compensated heat as a measure for "loss of potential work" has been emphasized by TOLMAN (see ref. 10 and patents cited there).

In EQUATION 4.1 and in all of the following equations to 4.16 we have dropped, for simplicity, the subscript j which should modify every symbol in these equations to specify that it refers to the j^{th} locality.

Here, T = temperature; p = pressure; γ = surface tension;

S = entropy; V = volume; n_i = moles of the i^{th} species;

m = mass; e = electric charge; σ = surface

$\mu_i = \frac{\partial E}{\partial n_i}$ = chemical potential of a mole of the i^{th} species

$\phi = \frac{\partial E}{\partial m}$ = gravitational potential of a gram of matter

$\psi = \frac{\partial E}{\partial e}$ = electrical potential of a coulomb of charge.

Now:

$$m_i = n_i M_i \quad \delta m_i = M_i \delta n_i$$

$$e_i = n_i Z_i \mathcal{F} \quad \delta e_i = Z_i \mathcal{F} \delta n_i$$

$$\sum_i \delta m_i = \delta m \quad \sum_i \delta e_i = \delta e$$

where M_i is the molecular weight, Z_i is the charge per molecule of the i^{th} species, and \mathcal{F} is the Faraday constant.

Accordingly (4.1) can be rewritten

$$\delta E = T\delta S - p\delta V + \gamma\delta\sigma + \sum_i (\mu_i + M_i\phi_i + Z_i\mathcal{F}\psi)\delta n_i. \quad (4.2)$$

Where the sum $(\mu_i + M_i\phi_i + Z_i\mathcal{F}\psi)$ can be conveniently replaced by the symbol λ_i , where λ_i is a general component potential for the i^{th} species; e.g. in the electrical case $(\mu_i + Z_i\mathcal{F}\psi)$ becomes the now well-known electrochemical potential of Guggenheim¹⁶ which Brønsted adopts and employs effectively in treating galvanic cells.^{1,8,9,17,18}

$T\delta S$ represents the heat which could be absorbed by the locality if the variation were conducted reversibly. In order to calculate δE , it is, therefore, demanded that the additional terms in (4.2) which represent work terms, be those which would obtain if the variation were conducted reversibly. In other words, p , γ , and λ_i must be equilibrium values. If we represent $\delta Q'$ as the heat which would be absorbed if the variation were conducted irreversibly, then

$$T\delta S - \delta Q' > 0. \quad (4.3)$$

Accordingly, we write

$$T\delta S = \delta Q' + T\delta S'' \quad (4.4)$$

where, by virtue of EQUATION 3:

$$T\delta S'' > 0 \quad (4.5)$$

in the irreversible case. In the reversible case, equality exists for (4.5) and the double primed quantity vanishes. The term $T\delta S''$ is the so-called non-compensated heat of Clausius represented by $\delta Q''$ while

$$\frac{\delta Q'}{T} = \delta S' \quad (4.6)$$

where $\delta S'$ is the entropy which is transported into the locality, through its boundaries. $\delta S''$ represents the entropy *produced within* the j^{th} locality by whatever irreversible phenomena are occurring there.

The variation of entropy, as ordinarily defined (no prime) is a sum given by

$$\delta S = \delta S' + \delta S'' = \frac{\delta Q'}{T} + \frac{\delta Q''}{T}. \quad (4.7)$$

For an irreversible variation in the j^{th} locality, the first law gives

$$\delta E = \delta Q' - \delta W' \quad (4.8)$$

where $\delta W'$ is the work performed by the locality upon its surroundings. Substituting for $\delta Q'$ we get

$$\delta E = T\delta S - T\delta S'' - \delta W' \quad (4.9)$$

EQUATION (4.9) indicates clearly that the non-compensated heat represents work which is potentially available, provided that the variation associated with δE is carried out reversibly. δE and δS have the fixed values, being exact differentials, independent of whether or not the change occurs reversibly.

Therefore, the sum of the residual terms

$$-T\delta S'' - \delta W' \quad (4.10)$$

is fixed for the defined variation.

In the limit of reversibility, $\delta S''$ is zero and consequently $\delta W'$ has its maximum value. All of the non-compensated heat can be obtained as useful work in this limit.

Now (4.10) is substituted into (4.9), yielding

$$\delta E = T\delta S' + T\delta S'' - p\delta V + \gamma\delta\sigma + \sum_i \lambda_i \delta n_i \quad (4.11)$$

for each j^{th} locality.

To compute the variation of the total internal energy of the isolated system we sum over two types of localities. The first summation is over all of the j localities. Frequently, the system may contain localities whose quantities are invariant to any general change. Fixed weights, or charges are examples. EQUATION (4.11) cannot be used for the computation of the

variation of the energy connected with the transport of such quantities, since all of its differentials are quantities and consequently equal to zero. Instead

$$\delta E_k = (\phi_k + \delta\phi_k - \phi_k)m_k = m_k\delta\phi_k \quad (4.12)$$

where ϕ is the gravitational potential and m is the mass is suitable if we deal with a weight, while

$$\delta E_k = (\psi_k + \delta\psi_k - \psi_k)\epsilon_k = \epsilon_k\delta\psi_k \quad (4.13)$$

is likewise suitable if we deal with an electric charge. Then the total variation of δE , the internal energy of the isolated system, is representable as

$$\delta E = \sum_j \delta E_j + \sum_k m_k \delta\phi_k + \sum_k \epsilon_k \delta\psi_k \quad (4.14)$$

and can be set equal to zero since the system is isolated. Accordingly, from (4.11) we obtain (4.15)

$$\begin{aligned} \delta E = \sum_j T_j \delta S'_j + \sum_j T_j \delta S''_j - \sum_j p_j \delta V_j + \sum_j \gamma_j \delta \sigma_j + \sum_j \sum_i \lambda_{ij} \delta n_{ij} \\ + \sum_k m_k \delta\phi_k + \sum_k \epsilon_k \delta\psi_k = 0. \end{aligned} \quad (4.15)$$

This equation can be rearranged immediately as follows:

$$\begin{aligned} -\sum_j T_j \delta S'_j + \sum_j p_j \delta V_j - \sum_j \gamma_j \delta \sigma_j - \sum_j \sum_i \lambda_{ij} \delta n_{ij} \\ - \sum_k m_k \delta\phi_k - \sum_k \epsilon_k \delta\psi_k = \sum_j T_j \delta S''_j. \end{aligned} \quad (4.16)$$

Since we have excluded the possibility of chemical reactions, all of the quantities on the left in (4.16) satisfy the condition of conservation for reversible processes in the isolated system. Thus:

$$\sum_j \delta S'_j = 0; \quad \sum_j \delta V_j = 0; \quad \sum_j \delta \sigma_j = 0; \quad \sum_j \delta n_{ij} = 0. \quad (4.17)^*$$

A further rearrangement of the left member of (4.16) can be effected in the following manner. Consider the sum $\sum_j p_j \delta V_j$ and specialize for simplicity to the case where it equals

$$p_1 \delta V_1 + p_2 \delta V_2 + p_3 \delta V_3 \quad (4.18)$$

Then the isolated system is a box, similar to that used in FIGURE 1, but having, in this case, (FIGURE 3) two movable partitions, separating regions having the pressures p_1 , p_2 , and p_3 , respectively. Now by the conservation of volume

$$\delta V_1 + \delta V_2 + \delta V_3 = 0. \quad (4.19)$$

* $\sum_j \delta S'_j = 0$ because $\delta S'_j$ is that part of the entropy which transported. $\sum_j \delta V_j$ and $\sum_j \delta \sigma_j$ can always be set equal to zero by defining, if necessary, transports to and from regions of zero pressure and zero surface tensions. Neither (4.15) (4.16) or (4.17) are affected by these transports because the terms referring to those localities of zero pressure and surface tensions necessarily have zero values. $\sum_j \delta n_{ij} = 0$ because we have excluded for the moment the possibility of chemical reactions.

It is apparent, from FIGURE 3, in which $P_1 > P_2 > P_3$, that

$$\begin{aligned}\delta V_2 &= \delta V_2' + \delta V_2'' \\ \delta V_1 &= -\delta V_2' + > 0 \\ \delta V_3 &= -\delta V_2'' > 0\end{aligned}\quad (4.20)$$

so that (4.19) is satisfied. Using the notation of (4.20) we can write for (4.18) the expression

$$p_1 \delta V_1 - p_2 \delta V_1 + p_2 \delta V_2'' - p_3 \delta V_2'' \quad (4.21)$$

or

$$(p_1 - p_2) \delta V_1 + (p_2 - p_3) \delta V_2''$$

In other words, (4.18) is identical with a sum, each term of which consists of the product of a potential difference, multiplied by the quantity which is transported through the potential difference. It is always easy to recognize the transported quantities.

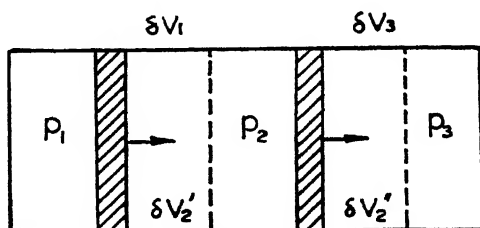


FIGURE 3.

By induction, this result is perfectly general, and can be applied to all of the sums on the left-hand side of (4.16). We thus effect the complete rearrangement, and write, finally, the important equation:

$$\delta A = \sum_r \delta K_r (P_{r(\text{initial})} - P_{r(\text{final})}) = \sum_j T_j \delta S_j'' \geq 0 \quad (4.22)$$

EQUATION (4.22) is the analytical form of the combined Brønsted principles. The equality applies to reversible processes, *i.e.*, the "work principle," and the inequality to irreversible processes, *i.e.*, the "equivalence principle." δK_r symbolizes a transported quantity, while P_r (initial) is the conjugate potential in the locality from which, and P_r (final) the conjugate potential in the locality to which, the quantity is transported. δK_r is always greater than (or equal to) zero—see EQUATION (4.20). The potential for volume always appears as negative pressure.

Virtual Changes, Coupling, and the Rule of Potentials

It is to be noted that all of the differentials in (4.22) are specified by the symbol, δ , which designates a virtual variation. The term, virtual, implies

that the variation may be of the most general kind, and need not be physically realizable. Although some of the most important phases of thermodynamic theory deal with the subject of virtual variations, very few authors have succeeded in presenting this subject clearly. The virtual variation is an important adjunct of the Brønsted treatment and hence requires discussion.

Any displacement from equilibrium subject to the same constraints which are imposed upon the system at equilibrium is, strictly speaking, impossible. A real displacement can only occur by the alteration of one or several of the constraints. In this sense, any displacement which a system in equilibrium undergoes is a virtual displacement. Let us confine our attention to displacements which are infinitesimal.

A system in equilibrium can experience two types of infinitesimal virtual displacements. The first type involves a displacement, joining equilibrium states (quasi-static), while the second involves a displacement which originates in equilibrium and terminates in non-equilibrium. The former is the asymptotic limit of some real process, while the latter has no basis in reality whatsoever.

The reader is undoubtedly familiar with many specific examples of quasi-static processes. Examples of this type will be given in the chapter on these processes.

For a concrete example of a non-quasi-static displacement, consider a liquid drop, in equilibrium, surrounded by its vapor. It would be impossible, unless the system were severely altered, to transport isothermally dn moles of the drop to the region of its vapor without, at the same time, transporting some of its volume and surface area. On the other hand, there is nothing to prevent us from imagining the physically impossible transport, during which the drop simultaneously dilates, so that its volume and surface area remain constant. It is evident that this would represent a displacement, passing from an equilibrium, to a non-equilibrium state.

In order to obtain a clear understanding of the usefulness of the method of virtual displacements, it is absolutely necessary to have a broader definition of a "thermodynamic state" than the one ordinarily given. Usually a "state" means an "*equilibrium state*" whose reproducible properties can be described by a minimum number of macroscopic parameters. Any function of state, *e.g.*, the free energy, has these parameters for arguments.

In a larger sense, a state can be defined as any reproducible condition of a system, either in equilibrium or in the process of change. A non-equilibrium state will, in general, require a larger number of parameters for its description, than an equilibrium state. In the extreme case, the dynamical specification of every microscopic particle in the system may be required. In any event, any function of state, *e.g.*, the free energy, will depend upon a larger number of variables, but will remain a defined function. From the operational point of view, an equilibrium state then becomes a special kind

of state, defined by the minimum number of parameters. It can be represented by a point in "state-space", *i.e.*, the space whose coordinates are the parameters defining the state, in the most general sense.

An infinitesimal displacement from equilibrium is represented by an infinitesimal path in "state-space" originating at the point of equilibrium. A number of these paths will satisfy the condition that the temperature and pressure remains constant along them. It is a classical criterion of equilibrium that, for an infinitesimal displacement along any one of these isothermal, isobaric paths, the free energy of the system remains unaltered. This free energy is understood to be defined in the larger sense, so that it remains a defined function of a non-equilibrium state. For the application of this criterion, it is inconsequential whether the displacement is or is not quasi-static. All that is demanded is that it be infinitesimal and that it originate in equilibrium. In particular, it may be of the type illustrated above, in connection with the spherical drop.

The point that many fail to grasp is that one does not seek information about the condition of the system along the path of the infinitesimal displacement but only about the condition at the origin of the path. Others have difficulty in conceiving the significance of the free energy along a non-quasi-static path because the description of a "state" as an equilibrium state has been over-emphasized.

The free energy has been chosen as an illustration because of its familiarity. However, all of these implications concerning virtual variations can be transferred, in full, to the "work principle," (4.22), when employed as the criterion of equilibrium. This equation (like other criteria) imposes the demands of thermodynamics upon a system in equilibrium. Very often, however, certain extra-thermodynamic conditions are imposed upon the behavior of the system. When this is true, all of the virtual displacements must be consistent with these conditions.

For example, return to the consideration of the drop. We may impose an extra-thermodynamic condition upon the system represented by the drop and its vapor, namely, the geometric condition which specifies that the transport of volume from the drop to the vapor must occur, in such a manner that the spherical shape of the drop is retained. It is then not permissible to carry out a virtual variation during which the volume, δv , is transported without the simultaneous transport of the surface, $\delta \sigma$, because both are connected by the geometrical relation

$$\delta v = \frac{r}{2} \delta \sigma \quad (5.1)$$

where r is the radius of the drop.

The equality and inequality (4.22) represents a compact and extremely useful expression of the laws of thermodynamics. In addition, it furnishes

a very satisfying model for the internal behavior of a thermodynamic system. These contentions shall be demonstrated in detail.

From the nature of the rearrangement (4.22), it is clear that the potentials conjugate to the different quantities are, in order, as shown in TABLE 1. In mechanics and field theory, "potential" has the significance of determining the direction of change. That this significance is retained, unaltered, in TABLE 1 can easily be shown.

To do this, consider a system undergoing a virtual change which consists of a single transport, such that all of the terms on the left of (4.22) with the exception of one, $\delta K_x(P_{x(\text{initial})} - P_{x(\text{final})})$, are zero. Then (4.22) reduces to

$$\delta A = \delta K_x(P_{x(\text{initial})} - P_{x(\text{final})}) = \sum_i T_i \delta S_i'' \geq 0 \quad (5.2)$$

In (5.2) as in (4.22) the inequality corresponds to a *natural* irreversible change, *i.e.*, one which does occur spontaneously, and the equality corresponds to a reversible process or to a displacement of a system in

TABLE 1

Quantity	Potential
volume	negative pressure
entropy	temperature
surface	surface tension
moles of chemical components	component potential*
mass	gravitational potential
charge	electrical potential

* Note that we are replacing the ordinary Gibbs chemical potential μ by the more general component potential defined in the chapter on the derivation of Brønsted's principles.

equilibrium. Since δK_x is arbitrary and positive, it follows that the expression in brackets (potential difference) is positively different from zero when a real change takes place. Finally, we observe that the potential difference is zero when no change takes place (when the system is in equilibrium). Therefore, a finite difference of potential bears a one-to-one correspondence to change, while no potential difference corresponds to no change. For this reason, potential difference may be regarded with complete consistency as the motivating factor for change. Taking account of the subscripts "initial" and "final" in (5.2), it is to be observed that all quantities tend to move from a higher to a lower potential. These conclusions which we have derived from the laws of thermodynamics, Brønsted introduces as observations of experience to justify the reasonableness of his principles.

The form (5.2) was achieved by restricting the virtual change to a single transport. But suppose this is not possible, as in the example offered previously, concerning the volume and area of a spherical drop. In that case the quantities, volume and area, were *coupled* together so that the movement of one demanded the movement of the other. For such a case, the form

(5.2) could not be achieved. Then it could not follow that the potential differences conjugate to the coupled quantities would be required to be zero at equilibrium.

We are thus led, quite rigorously, to a general rule which we shall call the *rule of potentials*, namely, *that all potential differences necessarily vanish at equilibrium except those corresponding to conjugate transported quantities which are coupled to other quantities*. In particular, since chemical components are never coupled so as to defy an *individual virtual transport*, the component potential λ_i corresponding to the i^{th} species is identical in every locality when equilibrium has been attained.

In the usual presentation of thermodynamics the rule of potentials, enunciated above, can only be proved by inventing a suitable characteristic function for each case and by setting in motion the machinery of the Lagrange method of undetermined multipliers. In the current presentation it has been obtained *rigorously* and in a *single stroke* by utilizing a satisfactory physical model for the thermodynamic system in which constraints can be described in terms of bonds, "coupling."

This result constitutes part of the evidence for the contention that (4.22) is a compact and useful expression of the laws of thermodynamics, and that it furnishes a good model of thermodynamic behavior. We shall now proceed to examine the beautiful and consistent description which it provides for the state of "internal equilibrium" when coupling exists.

A Model for Internal Equilibrium

When there is coupling, the potential differences conjugate to the coupled quantities are not necessarily zero. If we write (4.22) for the process involving the reversible transport of these quantities, we retain only the equality, and have:

$$\sum_{\text{coupled}} \delta K_r (P_{r(\text{initial})} - P_{r(\text{final})}) = 0. \quad (6.1)$$

Physically, the situation in (6.1) can be described as follows. Each of the coupled quantities is invited by its conjugate potential difference, to move. But the movement of one quantity, in the direction specified by its conjugate potential difference, compels (because of the bonds between quantities) other quantities to move in directions opposite to those specified by their own potential differences. At equilibrium, all of the opposing tendencies balance, and this is signified by the condition (6.1).*

Treatment of Quasi-static Processes

In the first place, it is to be noted that a quasi-static displacement is one along which the system remains in equilibrium. Consequently, all of the

* This model of equilibrium (*i.e.*, coupling between basic processes) was applied in Brønsted's last monograph⁹ to the treatment of the reversible aspects of steady state processes, *e.g.*, in the thermoelectric cell he utilizes the coupling between a mole of electrons and the entropy associated with it. Similar procedures were employed for the gas transpiration cell.

potentials in the system are subject to the restrictions of our rule of potentials. In addition, since a quasi-static displacement has a limit-basis in reality, we shall consider it formally to be a real change and accordingly employ the symbol d rather than δ to symbolize differentials.

THE REVERSIBLE EXPANSION OF A GAS. Consider a gas (FIGURE 4) having the pressure p , separated from a vacuum by a partition upon which a weight, m , rests which is almost, but not quite, heavy enough to maintain equilibrium. The containing vessel is surrounded by a reservoir of temperature, T . The isolated system which we need to consider consists of the reservoir and the container plus its contents.

Under the prescribed conditions, the partition will move upward the distance, dh , and the volume, dV , will be transported from the vacuum to the

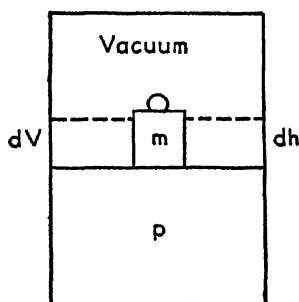


FIGURE 4.

pressure, p . Corresponding to this transport the loss of spatial potential work is given by

$$dA_{\text{spat}} = -(0 - p) dV = p dV. \quad (7.1)$$

There is an accompanying loss of gravitational potential work

$$dA_{\text{grav}} = [gh - g(h + dh)] m = -mg dh. \quad (7.2)$$

There will also be a flow of entropy between the reservoir and the gas. The variation is quasi-static and the rule of potentials can be applied. Since there is no extra-thermodynamic relation coupling the transport of entropy to other processes, it follows that the potentials conjugate to the entropy, *i.e.*, the temperatures are the same in the gas and in the reservoir. For the thermal loss of potential work, we obtain

$$dA_{\text{therm}} = 0. \quad (7.3)$$

However, there is an extra-thermodynamic, geometrical relation between dV and dh , and thus the spatial and gravitational processes *are* coupled. This is consistent with the rule of potentials; the corresponding losses of potential work do not have zero values.

Upon substitution of (7.1), (7.2), and (7.3) in the equality of (4.22), we obtain the result

$$p dV - mg dh = 0. \quad (7.4)$$

Brønsted has commented upon the fact that it is customary to regard the motivation of the quasi-static process just described as originating in the reversible flow of heat from the reservoir to the gas, or, what amounts to the same thing, in the transport of entropy. He points out that it is more consistent to regard the transport of volume as the motivating factor, since it is here that a finite potential difference exists (see EQUATION (7.1)), and we have considered a potential difference to be the motivating factor for change (see the section on virtual changes).

CARNOT CYCLE—COUPLING OF THERMAL AND MECHANICAL BASIC PROCESSES. As an important example of the direct use of the equality in (4.22) we give Brønsted's treatment of the reversible Carnot engine. Let the heat absorbed by the engine at the upper temperature T_1 be $DQ_1 = T_1 dS$ and that rejected at the lower temperature T_2 be $DQ_2 = T_2 dS$:

$$dS = \frac{DQ_1}{T_1} = \frac{DQ_2}{T_2}. \quad (7.5)$$

The thermal process in the heat engine thus consists in the reversible transport of an amount of entropy dS between the temperatures T_1 and T_2 . The reversible mechanical work, DA , obtained from the engine may consist in the transport of a weight m from a lower height h , to a higher height $h + dh$: $DA_{\text{mechanical}} = -mgdh$. In any case, regardless of the nature of the mechanical work, (4.22) gives:

$$(T_1 - T_2) dS + dA_{\text{mechanical}} = 0. \quad (7.6)$$

or, introducing (7.5):

$$DA_{\text{mechanical}} = DQ_1 \frac{T_1 - T_2}{T_1}. \quad (7.7)$$

To be consistent, the driving force may be considered to originate in the thermal process, *i.e.*, the tendency of the entropy to go from the higher to the lower temperature. Since this process is coupled and balanced through the engine to a mechanical process, it can only occur reversibly through the simultaneous performance of mechanical work.

The reader is referred to the original articles^{4,7} for Brønsted's objections to the Clausius' interpretation of the Carnot Cycle.*

The Gibbs-Duhem Equation

In treating problems of equilibrium, it is often necessary to have a differential relation which connects the real variations (symbol d) of the

* See also, V. K. LAMER, "Some current misconceptions of Carnot's Memoir and Cycle" Paper read before American Physical Society, Jan. 29, 1949, to be published in Am. J. Phys.

different potentials, rather than the extensive properties, obtaining in a given phase. The general relation, which we shall call the generalized Gibbs-Duhem equation, has the following form:

$$S dT - v dp + \sum_i n_i d\lambda_i = 0. \quad (8.1)$$

EQUATION (8.1) can be obtained, in a simple and straight-forward manner, by applying the equality contained in (4.22) to a selected reversible transport.

Consider a phase whose potentials are specified by the pressure p , the temperature T , and the component potential for the i^{th} species, λ_i . Consider another phase, having the potentials, $p + dp$, $T + dT$, and $\lambda_i + d\lambda_i$, which contains the same chemical species as the first phase. Now, combine these two phases in a rigid, adiabatic shell, so that they form an isolated system. Since the potentials in the two phases differ, infinitesimally, the transports which now occur, do so reversibly. We can thus apply the equality, contained in (4.22), to these transports.

$$[(T + dT) - T] dS + [p - (p + dp)] dV + \sum_i [(\lambda_i + d\lambda_i) - \lambda_i] dn_i = 0 \quad (8.2)$$

or

$$dS dT - dV dp + \sum_i dn_i d\lambda_i = 0. \quad (8.3)$$

This equation places no restriction upon the amounts of quantity dS , dV , and dn_i which are transported, since there is only one dependent variable, and we can always choose this to be one of the potential differences, *i.e.*, dT , dp , or $d\lambda_i$. By suitably adjusting the amounts of quantity, originally present in the two phases, it is always possible to adjust the transport so that:

$$dS : dV : dn_i = S : V : n_i. \quad (8.4)$$

Here, S , V , and n_i are the quantities in the first phase. This means that the quantities, transported, combine to form a replica of a portion of the first phase. Because of (8.4), (8.2) can be multiplied by a constant to yield (8.1), which is the Gibbs-Duhem equation for the first phase. Since the first phase was arbitrary, (8.1) is applicable to any phase.*

The derivation of (8.1) is again an illustration of the compactness and usefulness of (4.22). In the usual presentation of thermodynamics, it is necessary to invent a function of state, and to apply Euler's theorem for homogeneous functions, before (8.1) can be derived.

Treatment of Equilibrium

We are now in a position to apply (4.22) to the solution of problems of equilibrium. In a sense, we have already, partially solved every conceivable

* Our use of the equality (4.22) for the derivation of the generalized Gibbs-Duhem equation is slightly different from that of Björnsted.*

problem of internal equilibrium by the use of (4.22), since we have been able to arrive at the conclusion that the component potentials are uniform throughout the system when equilibrium has been attained. To obtain a more tangible and comprehensive description of the interior of a system at equilibrium, we have only to proceed from this point by the usual methods of thermodynamics, taking account of the manner in which the component potentials are related to the other parameters which determine the state of a given locality.

SURFACE TENSION AND INTERNAL PRESSURE OF A DROP. However, we have not exhausted the utility of the equality in (4.22) for, in many cases, it yields an immediately useful result, over and above that pertaining to the equality of the component potentials. As an example, compute the difference in pressures, inside and outside of a drop, having the radius, r . Choose, for the isolated system, the drop surrounded by its equilibrium vapor, contained in a rigid, diathermic shell which is placed in a thermostat. Since the temperature is, everywhere, uniform, the terms referring to the transport of entropy vanish from (4.22). The same, of course, is true of the transport of material. Writing the equality (4.22) for the transports attending the transport of moles of material from the drop to its vapor, we find that only the terms corresponding to the transport of the "coupled" quantities volume, δV , and surface $\delta\sigma$, can have non-zero values. This follows from the rule of potentials.

We thus have for (4.22):

$$-(p_1 - p_2)\delta V + (\gamma - 0)\delta\sigma = 0 \quad (9.1)$$

Here, p_2 is the pressure of the vapor, p_1 , the pressure inside the drop, 0, the surface tension of the hypothetical surface, in the vapor, and, γ , the surface tension of the vapor-drop interface. Substituting (5.1) into (9.1) the familiar formula of Kelvin, specialized to a sphere follows immediately.

$$p_1 - p_2 = \frac{2\gamma}{r}. \quad (9.2)$$

Chemical Equilibrium

Thus far, systems in which chemical reactions occur have been excluded from consideration. This was done as a matter of convenience only, and does not represent any fundamental insufficiency of the Brønsted treatment. The inclusion of the chemical reactions as a possible source of variation necessitates the introduction of a slight modification in equation (4.15).

The rearrangement of (4.15) to yield (4.17) is no longer valid, since any particular type of molecular species, entering into the reaction, is not conserved. It is possible to modify (4.15) so that in place of the mole numbers either the numbers of atoms of particular kinds, contained in a particular

molecular species, inhabiting a given phase, serve as parameters. It is also possible to use the masses of the various molecular species in this connection. Both atoms and mass are conserved, even in the presence of a chemical reaction, and so a rearrangement of the desired type is possible.

However, it is more expedient, for chemical purposes, to define a pseudo-quantity $\delta\alpha$, which is also conserved. Let ν_r and ν_p be the stoichiometric coefficients of the r^{th} reactant and p^{th} product, in a given chemical reaction of the type: $\sum \nu_r R_r = \sum \nu_p P_p$, where R_r and P_p are the molecular symbols of the r^{th} reactant and p^{th} product. Let ν_r and ν_p both be positive. Then*

$$\delta\alpha_R = \frac{\delta n_r}{\nu_r} \quad \text{for all } r \quad (10.1)$$

$$\delta\alpha_P = \frac{\delta n_p}{\nu_p} \quad \text{for all } p. \quad (10.2)$$

From stoichiometric considerations, it is evident that

$$\delta\alpha_P + \delta\alpha_R = 0. \quad (10.3)$$

For simplicity let us restrict our attention to a system in which the mole numbers are varied by a single chemical reaction confined to a single phase. This result can be generalized easily, as the occasion requires.

Then the term in (4.18), $-\sum \lambda_i \delta n_i$, reduces to $-\sum \lambda_i \delta n_i$, and by virtue of (10.1) and (10.2) this becomes

$$-(\sum_r \nu_r \lambda_r \delta\alpha_R + \sum_p \nu_p \lambda_p \delta\alpha_P) \quad (10.4)$$

or

$$-(\delta\alpha_R \sum_r \nu_r \lambda_r + \delta\alpha_P \sum_p \nu_p \lambda_p) \quad (10.5)$$

and by the use of the new conservation condition (10.3), we obtain the form

$$(\sum_r \nu_r \lambda_r - \sum_p \nu_p \lambda_p) \delta\alpha_P. \quad (10.6)$$

If we define

$$\lambda_P = \sum_p \nu_p \lambda_p \quad (10.7)$$

$$\lambda_R = \sum_r \nu_r \lambda_r \quad (10.8)$$

as "system potentials" for the pseudo-quantity, $\delta\alpha_P$, (10.6) indicates that the form (4.22) can be extended to chemical reactions.

Indeed, for any reaction proceeding isothermally and isobarically, the work principle now demands that at equilibrium

$$(\lambda_R - \lambda_P) \delta\alpha_P = 0 \quad (10.9)$$

or that the "system potentials"

$$\lambda_R = \lambda_P \quad (10.10)$$

* It will be noted that α_P , but not α_R , is the degree of advancement of reaction employed by De Donder.

(10.10) yields the law of mass action when the individual potentials are substituted.

In closing, it is to be noted that the pseudo-quantity can be used as a measure (on stoichiometric grounds) of the rate of transport of the real quantity, mass, from reactants to products.

Inversible Processes

The inequality contained in (4.22) provides a direct means for computing the production of non-compensated heat during an irreversible process, provided that the transports involved are recognizable and that the irreversible process conducts itself in such a way that each stage can be described by what are sensibly equilibrium parameters.

We shall consider one example of this type. A single thermostatted phase, the seat of a chemical reaction, but nevertheless in *mechanical* and *thermal* equilibrium, represents a system satisfying the requirements just mentioned. The only transport having a non-zero term will be that corresponding to the transport of the pseudo-quantity $d\alpha_P$. The system is out of equilibrium so that entropy is being produced. EQUATION 4.22 then reduces to

$$(\lambda_R - \lambda_P)d\alpha_P = TdS''. \quad (11.1)$$

If we divide by dt and define the velocity of reaction, v , as

$$v = \frac{d\alpha_P}{dt} \quad (11.2)$$

we obtain

$$(\lambda_R - \lambda_P)v = T \frac{dS''}{dt} \quad (11.3)$$

or

$$\frac{dS''}{dt} = \frac{(\lambda_R - \lambda_P)}{T} v \quad (11.4)$$

a result given by De Donder.

Assessment

The favorable points for EQUATION (4.22) follow:

(a) It provides a satisfying model for the internal behavior of an isolated thermodynamic system.

(b) It leads simply, and with a minimum of mathematical expenditure, to a simple rule of potentials. As a corollary, the general result asserting that the component potentials are uniform, at equilibrium, is obtained.

In the classical discipline, the concepts and ideas are not available to make such a concise universal statement.

(c) The generalized Gibbs-Duhem relation is obtained with a minimum of mathematical expenditure.

(d) In cases of coupling, (4.22) leads to an immediately useful result concerning the features of equilibrium in a system where the coupling phenomena exist. By this it is implied that properties, other than the fact that the component potentials are uniform, are described.

(e) In some instances, (4.21) affords a direct means of calculating the non-compensated heat evolved in an irreversible change.

Finally, we do not assert that EQUATION 4.22 is the most convenient form for all thermodynamic purposes. Attention is always focused upon an isolated system, which means any system of physical interest *plus* its environment. In this way, some of the detachment which is gained by defining thermodynamic potentials which are functions of the state of some particular non-isolated system is lost. However, by combining both methods of attack, fruitful results are obtained.

Summary

A brief exposition of the salient features of Brønsted's Energetics is given. The complete equivalence of his basic postulates, namely, the work and the heat and equivalence principles in respect to the two laws of classical thermodynamics, has been demonstrated by deriving his postulates from these laws. Some of Brønsted's fundamental conceptions, *e.g.*, the existence of a potential difference as the motivating factor for the occurrence of a basic process, balanced coupling of basic processes to produce reversible processes, the localized production of entropy in irreversible processes, *etc.*, emerge as necessary consequences in this derivation.

The compactness and elegance of Brønsted's approach are illustrated by simple examples using his work principle and a rule of potentials given by us. An assessment of the merits of the system is given.

Bibliography

1. BRØNSTED, J. N. 1937. *Physical Chemistry*. Heinemann, London; Chemical Publishing Co., New York. (Translated from the 1st Danish Ed. 1936 by R. P. Bell.)
2. BRØNSTED, J. N. 1937. D. Kgl. Danske Vid. Selsk. Medd. 15 (4).
3. BRØNSTED, J. N. 1939. D. Kgl. Danske Vid. Selsk. Medd. 16 (10).
4. BRØNSTED, J. N. 1940. Phil. Mag. 7 (29): 449.
5. BRØNSTED, J. N. 1940. J. Phys. Chem. 44: 699.
6. MACDOUGALL, F. H. 1940. J. Phys. Chem. 44: 713.
7. BRØNSTED, J. N. 1941. *On the Concept of Heat* (in English). D. Kgl. Danske Vid. Selsk. Medd. 19 (8): 39, 41.
8. BRØNSTED, J. N. 1943. *Fysisk Kemi*, 2nd ed. Munksgaard. Copenhagen.
9. BRØNSTED, J. N. 1946. *Principer og Problemer i Energetikken*. Københavns Univ. Festschrift. Bianco Luno. Copenhagen.
10. TOLMAN, R. C., & P. C. FINE. 1948. Rev. Mod. Phys. 20: 51.
11. ECKART, C. 1940. Phys. Rev. 58: 267, 269, 919.

12. BRIDGMAN, P. W. 1940. Phys. Rev. **58**: 845. 1941. *The Nature of Thermodynamics*, 133-147. Harvard Univ. Press.
13. DE DONDER, TH., & P. VAN RYSSELBERGHE. 1936. *Affinity*. esp. p. 9. Stanford Univ. Press.
14. PRIGOGINE, I. 1947. *Etude Thermodynamique des Phénomènes irréversibles*: esp. p. 1 and 2. Desoer. Liège.
15. LEAF, B. 1944. J. Chem. Phys. **12**: 89.
16. GUGGENHEIM, E. A. 1929. J. Chem. Phys. **33**: 842. 1936. *Modern Thermodynamics*, chap. 10. Methuen. London.
17. BRØNSTED, J. N. 1929. Z. Phys. Chem. **A143**: 301.
18. BRØNSTED, J. N. 1933. *On the Definition of the Gibbs Potential* (in English). D. Kgl. Danske Vid. Medd. **12**: 6.

THE EFFECTS OF SHAPE ON THE INTERACTION OF COLLOIDAL PARTICLES

BY LARS ONSAGER

Sterling Chemistry Laboratory, Yale University, New Haven, Connecticut

Introduction. The shapes of colloidal particles are often reasonably compact, so that no diameter greatly exceeds the cube root of the volume of the particle. On the other hand, we know many colloids whose particles are greatly extended into sheets (bentonite), rods (tobacco virus), or flexible chains (myosin, various linear polymers).

In some instances, at least, solutions of such highly anisometric particles are known to exhibit remarkably great deviations from Raoult's law, even to the extent that an anisotropic phase may separate from a solution in which the particles themselves occupy but one or two per cent of the total volume (tobacco virus, bentonite). We shall show in what follows how such results may arise from electrostatic repulsion between highly anisometric particles.

Most colloids in aqueous solution owe their stability more or less to electric charges, so that each particle will repel others before they come into actual contact, and effectively claim for itself a greater volume than what it actually occupies. Thus, we can understand that colloids in general are apt to exhibit considerable deviations from Raoult's law and that crystalline phases retaining a fair proportion of solvent may separate from concentrated solutions. However, if we tentatively increase the known size of the particles by the known range of the electric forces and multiply the resulting volume by four in order to compute the effective van der Waal's co-volume, we have not nearly enough to explain why a solution of 2 per cent tobacco virus in 0.005 normal NaCl forms two phases.

General Kinetic Theory and Conventions. Some care is needed when we apply the general principles of statistical thermodynamics to solutions of colloidal particles. On one hand, any force acting on a particle of whatever size is important as soon as the work of the force is comparable to kT . On the other hand, the presence of one colloidal particle will usually affect the free energy of dilution of the electrolyte present by a large multiple of kT . This difficulty must be circumvented by all theories and experiments pertaining to the distribution of colloidal particles. One suitable piece of experimental apparatus is an osmometer whose membrane is impermeable to the colloidal particles, but permeable to all small molecules and ions of the electrolytic solvent. The osmotic pressure measured across such a membrane will be exactly proportional to the number of particles if the solution behaves like an ideal gas. The analogy can be extended to real gases and real solutions, whereby the gas pressure still corresponds to osmotic pressure.

The imperfection of an ideal gas can be computed when we know the forces between the molecules for every configuration. For that purpose, we have to evaluate the integral

$$B(T) = \int e^{-u/kT} d\tau / N! \quad (1)$$

where u stands for the potential of the forces and $d\tau$ denotes a volume element in configuration-space. The free energy of the gas in terms of this integral is

$$F(N, V, T) = N\mu_0(T) - kT \log B(N, V, T) \quad (2)$$

where the additional function $\mu_0(T)$ depends only on the temperature and does not enter into the computation of the pressure,

$$P = -(\partial F / \partial V)_{N,T} = kT(\partial \log B / \partial V). \quad (3)$$

The osmotic properties of a colloidal solution can be computed by a similar procedure. What we need to know initially is the *potential*

$$w((q_1), (q_2), \dots (q_N))$$

of the *average forces* which act between the particles in a configuration described by the sets of coordinates $(q_1), (q_2), \dots (q_{N_p})$ of particles 1, 2, $\dots N_p$. In general, it is necessary to specify the orientations of the particles as well as the positions of their centers, and the work against the corresponding torques must be included in w .

With

$$B_p(N_p, V, T) = \int e^{-w/kT} d\tau / N_p! \quad (4)$$

we have then

$$\begin{aligned} & F(\text{solution}) - F(\text{solvent}) \\ &= N_p \mu_p^\circ(T, \text{solvent}) - kT \log B_p(N_p, V, T). \end{aligned} \quad (5)$$

Here, the difference between "solution" and "solvent" means that the former contains colloidal particles, and we compare solutions of different colloid concentrations $c = (N_p/V)$ always in "dialytic" equilibrium across an osmometer membrane with a "solvent" of constant composition. The proportions of ions and molecules present between the particles in the colloidal solution may differ from those in the "solvent" as we have defined it. This complication can hardly be avoided if we want simple relations and *precise interpretation of practicable experiments*.

With these conventions the analog of EQUATION 3,

$$P = kT(\partial \log B_p / \partial V), \quad (6)$$

is valid for the osmotic pressure and

$$\mu_P = \mu_P^0 - kT(\partial \log B_P / \partial N_P) \quad (7)$$

for the chemical potential of the colloidal particles.

Moreover, the conditions for coexistence of two phases are simply

$$P = P' \quad (8a)$$

$$\mu_P = \mu_P' \quad (8b)$$

The assumed dialytic equilibrium takes care of all small molecules and ions.

Electric Forces. According to theories developed by Helmholtz, Lamb, and Smoluchowski, the speed of migration of a colloid in an electric field is quantitatively related to the potential difference between the first mobile layer of liquid in contact with the particle and the bulk of the solution. It is customary to specify the electric charges of particles indirectly in terms of this so-called ζ -potential. The theory is still somewhat incomplete as regards cases where the thickness of the electric double layer is of the same order of magnitude as the dimensions of the particle; a factor variable between the limits of unity and 3/2 then enters into the interpretation.

With slight approximations, the general kinetic theory for the distribution of ions near charged particles leads to the well-known Poisson-Boltzmann differential equation for the electric potential

$$\nabla^2 \psi \equiv \frac{\partial^2 \psi}{\partial x^2} + \frac{\partial^2 \psi}{\partial y^2} + \frac{\partial^2 \psi}{\partial z^2} = - \frac{4\pi}{D} \sum_i n_i e_i e^{-e_i \psi / kT} \quad (9)$$

where e_1, e_2, \dots denote the charges of ions present in concentrations n_1, n_2, \dots (in the solution or, rather, in a "solvent" maintained in dialytic equilibrium), and D denotes the dielectric constant. Whenever the condition $|e_i \psi| \ll kT$ is satisfied for all kinds of ions present, EQUATION 9 may be replaced by

$$\begin{aligned} \nabla^2 \psi &= \kappa^2 \psi \\ \kappa^2 &= (4\pi / DkT) \sum_i n_i e_i^2 \end{aligned} \quad (10)$$

The normal gradient of ψ at the surface of the particle is related to the charge density on the particle. We have to expect an implicit boundary condition determined by the adsorption and surface ionization in equilibrium with ions present at the surface in local concentrations $n_1 \exp(-e_1 \zeta / kT)$, $n_2 \exp(-e_2 \zeta / kT)$, etc. Since the kinetics of the surface ionization is rarely known, the relation

$$\psi = \zeta = \text{constant}; \quad (\text{at surface}) \quad (11)$$

has often been assumed, regardless of modifying factors, although systematic variations of ζ with electrolyte concentration, etc., should be expected and have been demonstrated in some cases. We shall not pursue these questions, because the expected variations of ζ will have but little effect on the forces between the particles.

In one-dimensional cases, $\psi = \psi(x)$, EQUATION 9 is generally soluble by quadratures. However, even the simplest case of a binary electrolyte between parallel plates, both maintained at the potential ζ , leads to elliptic integrals (Langmuir, 1938), and the resulting exact formula for the force is fairly involved (Verwey and Overbeek, 1948). Their approximation

$$K(x) = 16 n kT (\tanh(e_1 \zeta / 4kT))^2 e^{-\kappa x} \quad (12)$$

for the force per unit area between two parallel plates separated by a distance x is valid for not too small distances and will suffice as a basis for discussion. We note that the force decreases exponentially and that the screening constant κ , given by EQUATION 10, depends only on the ionic strength of the solvent. Moreover, for any fixed distance d between the plates, the force approaches a finite limit with increasing particle potential ζ . These two features are general.

We may use the result of EQUATION 12 to estimate the force between two infinite cylinders of the same diameter d crossing at an angle γ in such a manner that the mantles are separated by a distance x_0 between the points of closest approach.

We choose Cartesian coordinates in a plane parallel to the axes of both cylinders and identify points on the cylinder mantles by the coordinates (y, z) of their projections upon that plane. Then, the distance between points on the two cylinder mantles with the same (y, z) coordinates will be:

$$x(y, z) = x_0 + d - (\tfrac{1}{4} d^2 - y^2)^{1/2} - [\tfrac{1}{4} d^2 - (y \cos \gamma - z \sin \gamma)^2]^{1/2}.$$

If we allow the approximation

$$x(y, z) \sim x_0 + (y^2/d) + ((y \cos \gamma - z \sin \gamma)^2/d),$$

and compute the local force density $K(x)$ according to EQUATION 12, an elementary integration yields for the total force

$$\text{Average Force} = (\pi d / \kappa \sin \gamma) K(x_0) \quad (13)$$

and we obtain for the potential w of the average force

$$w/kT = (d/q \sin \gamma) [\tanh(e_1 \zeta / 4kT)]^2 e^{-\kappa x_0} \quad (14)$$

where we use the abbreviation

$$q = e_1^2 / 2 D kT = z_1^2 \times 3.56 \times 10^{-8} \text{ cm.} \quad (15)$$

and $e_1 = -e_2$ denote the charges, $z_1 = -z_2$ the valences of the ions in the solvent; the numerical value refers to water at 25°C.

As an example, we may consider two perpendicular cylindrical particles of $d = 150$ A. U., $\zeta = 0.15$ volt, in a 0.005 mol NaCl solution, whereby $1/\kappa = 43$ A. U. We find

$$w/kT = 34.5 e^{-\kappa x_0} \quad (16)$$

which equals $e^{-c} = 0.561$ at a distance $x_0 = \delta = 4.12/\kappa = 184$ A. U.

Here, we have neglected the divergence of the electric force-lines, which must be quite appreciable because δ is by no means small, compared to d . We apply a correction of the right order of magnitude if we multiply w by the factor $d/(d + x_0)$; according to the corrected formula we then find $w = 0.561 kT$ at a distance of about 151 A.U.

It will be evident that over a considerable range of particle diameters and orientations and over a wide range of concentrations of electrolyte, the effective range of the electrostatic repulsion will be a modest multiple of the screening distance $1/\kappa$. While exact computations are not available, there can be little doubt about the orders of magnitude involved.

One further observation is in order: unless the electric double layers of three particles overlap *in the same region*, the repulsive forces are additive. When $\kappa d \gg 1$, the exceptional configurations are just about impossible; but even under much less stringent conditions very few of them can occur. On the strength of these estimates, we shall treat the electrostatic repulsion as an additive *short range effect*. For very low concentrations of ions such that κd is small, our procedure may be unreliable. On the other hand, we shall make no allowance for differential van der Waal's attraction. This omission would tend to become particularly serious for high concentrations of ions and low ζ -potentials, under conditions approaching those which lead to flocculation of the particles.

Imperfect Gas Theory. We proceed to evaluate the configuration integral of Equation 4 according to the general method developed by Mayer and Mayer. Assuming additive forces:

$$w = w_N((q_1), \dots (q_N)) = \sum_{i < j} w_{ij}; \quad (17)$$

$$w_{ij} = w_2((q_i), (q_j))$$

we put

$$\Phi_{ij} = \Phi_{ij}((q_i), (q_j)) = e^{-w_{ij}/kT} - 1. \quad (18)$$

In order to avoid confusion with a distribution-function f , we write Φ_{ij} for the functions which Mayer and Mayer denote by f_{ij} . Their notation sometimes implies the hypothesis that w_{ij} , and with it Φ_{ij} , depends only on the distance between two particles. Their specialization is not essential and their method is valid, with obvious pertinent modifications, for the more general case with which we have to deal.

Upon suitable rearrangement of the sum

$$e^{-w/kT} = 1 + \sum_{i>j} \Phi_{ij} + \sum \Phi_{ij} \Phi_{i'j'} + \dots \quad (19)$$

which now constitutes the integrand of EQUATION 4, Mayer and Mayer obtain an expansion for the integral in terms of the irreducible cluster integrals

$$\beta_1 = \frac{1}{V} \int \Phi_{12} d\tau_1 d\tau_2 \quad (20)$$

$$\beta_2 = \frac{1}{2V} \int \Phi_{12} \Phi_{23} \Phi_{31} d\tau_1 d\tau_2 d\tau_3$$

and these furnish the first two correction terms to the ideal gas laws in the expansion

$$\log B_p = N_p \{ 1 + \log (V/N_p) + \frac{1}{2} \beta_1 (N_p/V) + \frac{1}{2} \beta_2 (N_p/V)^2 + \dots \}. \quad (21)$$

Similarly, for a solution which contains N_1, \dots, N_s, \dots particles, of different types $1, \dots, s, \dots$, respectively, we have

$$\log B_p = \sum_s N_s (1 + \log (V/N_s)) + \frac{1}{2V} \sum_{s,s'} \beta_1(s, s') N_s N_{s'}$$

$$+ \frac{1}{3V^2} \sum_{s,s',s''} \beta_2(s, s', s'') N_s N_{s'} N_{s''} + \dots \quad (22)$$

The arguments of the cluster integrals indicate that the functions Φ_{12} , Φ_{23} , \dots involve the interaction potentials w appropriate to pairs of particles from the sets of types (s, s') , (s, s', s'') , etc.

In EQUATION 21, the generalized volume elements $d\tau$, are ordinary volume elements whenever the forces are central, so that w , and with it Φ , depend only on the distance between the two particles involved; but we shall be very much interested in the mutual orientations of the particles. In dealing with *isotropic* solutions, we have two alternative procedures at our disposal. The first method is to include an averaging over orientations (Ω) in the definitions of volume elements, thus

$$d\tau_j = dV_j d\Omega_j / \int d\Omega_j. \quad (23)$$

With particles of axial symmetry, it is, of course, enough to specify the directions of the symmetry axes, so that, for a cylindrical particle, we may let $d\Omega$ be an element of solid angle including the direction \mathbf{a} , of the cylinder axis:

$$d\tau_j = dV_j d\Omega_j / 4\pi. \quad (23a)$$

The second method is more general, in that it applies to anisotropic phases without periodic structure, in other words, liquid crystals of the *nematic*

type. For the purpose of computing B_p , we then introduce the artifice that we treat particles of different orientation as particles of different kinds. The distribution of particles among different orientations is determined by the condition that B_p must be a maximum. Incidentally, the convention that the terms $\log (V/N)$ are now formed separately for each "kind" of particles makes due allowance for the entropy of "mixing" (Gibbs Paradox). On the other hand, we must remember that the generalized volume in space and orientation available to a particle of orientation restricted to an element of solid angle $d\Omega$ is only $Vd\Omega$, rather than $4\pi V$ for a particle of unrestricted orientation. Thus, when we divide the total of all directions in space among elements of solid angle $\Delta\Omega_1, \dots, \Delta\Omega_\nu, \dots, \Delta\Omega_s$ surrounding the directions $\mathbf{a}_1, \dots, \mathbf{a}_\nu, \dots, \mathbf{a}_s$, respectively, these will have populations of particles which we shall denote by

$$\Delta N_\nu = N_\nu f(\mathbf{a}_\nu) \Delta\Omega_\nu; \nu = 1, 2, \dots, s, \quad (24)$$

whereby, of course,

$$\sum_1^s \Delta N_\nu = N_p \sum_\nu f(\mathbf{a}_\nu) \Delta\Omega_\nu = N_p. \quad (24a)$$

With this notation, EQUATION 21 is generalized as follows:

$$\begin{aligned} \log B_p = & \sum_\nu \Delta N_\nu (1 + \log (V \Delta\Omega_\nu / 4\pi \Delta N_\nu)) \\ & + \frac{1}{2V} \sum_{\nu, \nu'} \beta_1(\mathbf{a}_\nu, \mathbf{a}_{\nu'}) \Delta N_\nu \Delta N_{\nu'} \\ & + \frac{1}{3V^2} \sum_{\nu, \nu', \nu''} \beta_2(\mathbf{a}_\nu, \mathbf{a}_{\nu'}, \mathbf{a}_{\nu''}) \Delta N_\nu \Delta N_{\nu'} \Delta N_{\nu''} + \dots \end{aligned} \quad (25)$$

Here, the cluster integrals of EQUATION 20 are computed for fixed orientations $\mathbf{a}_1, \mathbf{a}_2$, viz. $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3$, of the particles involved. Replacing the sums by integrals in terms of the distribution-function $f(\mathbf{a})$, the integral of which is now normalized:

$$\int f(\mathbf{a}) d\Omega(\mathbf{a}) = 1, \quad (26)$$

we may write EQUATION 23 in the form

$$\begin{aligned} \log B_p = & N_p \left\{ 1 + \log (V/N_p) - \int f(\mathbf{a}) \log (4\pi f(\mathbf{a})) d\Omega(\mathbf{a}) \right. \\ & + (N_p/2V) \iint \beta_1(\mathbf{a}, \mathbf{a}') f(\mathbf{a}) f(\mathbf{a}') d\Omega d\Omega' \\ & \left. + (N_p^2/3V^2) \iiint \beta_2(\mathbf{a}, \mathbf{a}', \mathbf{a}'') f(\mathbf{a}) f(\mathbf{a}') f(\mathbf{a}'') d\Omega d\Omega' d\Omega'' + \dots \right\}. \end{aligned} \quad (27)$$

The Cluster Integrals. When the forces are repulsive at all distances, we have $w \geq 0$ everywhere, whence the functions Φ_{ij} , defined by EQUATION 19, satisfy the inequalities

$$-1 \leq \Phi_{ij} \leq 0 \quad (28)$$

everywhere. In this case, the first two cluster integrals β_1 and β_2 , defined by EQUATION 21, are necessarily negative, because the integrands are formed from one and three negative factors, respectively. The former has a particularly simple geometrical meaning in the ideal case of "hard" particles, which repel each other at contact but do not interact otherwise.

In this case, we have

$$\begin{aligned} w_{ij} &= +\infty; & \Phi_{ij} &= -1; & (\text{particles intersecting}) \\ w_{ij} &= 0; & \Phi_{ij} &= 0; & \text{otherwise} \end{aligned}$$

and $(-\beta_1)$ then equals the volume which is denied to particle j by the condition that it must not intersect particle i . For a pair of spheres of radius r , the excluded volume is obviously a sphere of radius $2r$. This leads to the familiar result first derived by Boltzmann, that the van der Waal's "co-volume" (per particle) equals four times the volume of one spherical particle

$$b = -\frac{1}{2}\beta_1 = 4(4\pi r^3/3) = 4v_p$$

The analogous problem for two cylinders of lengths of l_1, l_2 and diameters d_1, d_2 is solved in the Appendix; we reproduce here the result (from A 11)

$$\begin{aligned} -\beta_1(\gamma) &= (\pi/4)d_1 d_2 (d_1 + d_2) \sin \gamma \\ &+ (\pi/4)(l_1 d_1^2 + l_2 d_2^2) + (\pi/4)(l_1 d_2^2 + l_2 d_1^2) |\cos \gamma| \\ &+ (l_1 + l_2)d_1 d_2 E(\sin \gamma) + l_1 l_2 (d_1 + d_2) \sin \gamma, \end{aligned} \quad (30)$$

where $E(\sin \gamma)$ denotes the complete elliptic integral of the second kind

$$E(\sin \gamma) = \int_0^{\pi/2} (1 - \sin^2 \gamma \sin^2 \phi)^{1/2} d\phi. \quad (30a)$$

For special orientations or dimensions, the formula simplifies more or less. The following cases are instructive:

$$\gamma = 0: \quad -\beta_1 = (\pi/4)(l_1 + l_2)(d_1 + d_2)^2 \quad (31a)$$

$$\gamma = 0: \quad l_1 = l_2 = l; \quad d_1 = d_2 = d: \quad 8(\pi/4)l d^2 \quad (b)$$

$$\gamma = \pi/2: \quad l_1 l_2 (d_1 + d_2) + (l_1 + l_2 + d_1 + d_2)d_1 d_2 + (\pi/4)(l_1 d_1^2 + l_2 d_2^2)$$

$$l_1 = l_2; \quad d_1 = d_2: \quad (c)$$

$$(2 l d + (\pi/2)d^2) \sin \gamma + \{(\pi/2)(1 + |\cos \gamma|) + 2 E(\sin \gamma)\} l d^2, \quad (d)$$

$$l_1 = l_2 = 0: \quad (\pi/4)d_1 d_2 (d_1 + d_2) \sin \gamma \quad (e)$$

$$l_1 = d_2 = 0: \quad (\pi/4)l_2 d_1^2 |\cos \gamma| \quad (f)$$

$$l_2 = d_2 = 0: \quad (\pi/4)l_1 d_1^2 \quad (g)$$

$$l_1 \gg d_1 + d_2 \ll l_2: \quad l_1 l_2 (d_1 + d_2) \sin \gamma. \quad (h)$$

Case (b) yields 8 times the volume of one particle, as for spheres. This is generally true for centrosymmetrical convex particles in parallel orientation. Most of the others explain themselves. We call attention to the idealized cases (e) and (f), where the particles have a mutual covolume although neither has any volume, and to the case (h), which shows that the ratio (covolume/volume) for long needles is (length/diameter) rather than (4/1).

The theory of isotropic solutions involves a simple average of EQUATION 30 over all directions in space:

$$\begin{aligned}
 & -\bar{\beta}_1(l_1, d_1; l_2, d_2) \\
 & = 2b = - \int \beta_1(\gamma_{12}) d\Omega_{12}/4\pi \\
 & = -\frac{1}{2} \int \beta_1(\gamma) \sin \gamma d\gamma \\
 & = (\pi/4)^2 d_1 d_2 (d_1 + d_2) + (\pi/4)(l_1 d_1^2 + l_2 d_2^2) \\
 & + (\pi/8)(l_1 d_2^2 + l_2 d_1^2) + (\pi^2/8)(l_1 + l_2)d_1 d_2 \\
 & + (\pi/4)l_1 l_2 (d_1 + d_2). \tag{32}
 \end{aligned}$$

For details of the integration, we again refer to the Appendix (A 14). For particles of equal diameters $d_1 = d_2 = d$, EQUATION 32 simplifies

$$\begin{aligned}
 -\bar{\beta}_1(l_1, l_2) & = 2b_{12} = \frac{1}{2} \pi d \{l_1 l_2 + \frac{1}{4}(\pi + 3)(l_1 + l_2) d + \frac{1}{4} \pi d^2\} \tag{33} \\
 & = 1.5708 d(l_1 l_2 + 1.5354(l_1 + l_2) d + 0.7854 d^2),
 \end{aligned}$$

and, when the lengths as well as the diameters are equal, it simplifies still a little further

$$-\bar{\beta}_1 = \frac{1}{2} \pi d(l^2 + \frac{1}{2}(\pi + 3)l d + \frac{1}{4} \pi d^2). \tag{34}$$

It is interesting to examine the ratio of covolume to volume as a function of the ratio (l/d) according to EQUATION 34. The ratio

$$-\bar{\beta}_1/2(\pi/4) d^2 l = b/(\pi/4) d^2 l = b/v_p$$

becomes in various limiting cases

$$b/v_p \sim l/d; \quad (l \gg d)$$

$$b/v_p = \text{minimum} = \pi^{1/2} + \frac{1}{2}(\pi + 3) = 4.843; \quad (l = (\pi/4)^{1/2} d)$$

$$b/v_p \sim (\pi/4) d/l; \quad (l \ll d).$$

When the dimensions are about equal the ratio is not much more than 4, but for highly anisometric particles, whether needles or pancakes, (b/v_p) is about equal to the ratio of the long to the short dimension.

While the evaluation of the first cluster integral β_1 defined by EQUATION 20 proved perfectly feasible, the integral β_2 depends on three directions, and to

compute it exactly would be an extremely tedious task at best. For that reason we shall be content to estimate the order of magnitude of β_2 .

The value for spheres of equal diameters was computed by Boltzmann; in that case, one finds

$$-\beta_2 = (15/64)\beta_1^2.$$

This result gives us the right order of magnitude of the ratio (β_2/β_1) for isometric particles in general. Where anisometric particles are concerned, we must distinguish between slender rods and thin plates. For the latter case, a little experimentation with various orientations will show that in most cases where two plates intersect each other, the volume within which a third plate of comparable diameter will intersect the other two simultaneously will be a sizable fraction of the volume within which it will intersect a given one of the others. Accordingly, barring special orientation, the have the result

$$\beta_2/\beta_1 = O(-\beta_1); \text{ (spheres, cubes, plates).}$$

For the slender rods, we obtain the same result only if the three rods are nearly coplanar, whereby the admissible deviation in angle is of the order (d/l) . Otherwise, it is easily seen that if we look at a pair of intersecting rods along the direction of a third, the projection of their intersection upon a plane normal to the axis of the third rod will be (at most):

$$d_1 d_2 / \sin \phi_3,$$

where ϕ_3 is the angle between the projections of two rods, alias the angle between the planes containing the pairs of directions (a_1, a_3) and (a_2, a_3) , respectively. Or, considering the spherical triangle whose corners have the directions a_1, a_2, a_3 , ϕ_3 is the angle at the third corner. If we denote the three sides of this triangle (angles between the directions pairwise) by γ_{12}, γ_{23} and γ_{31} , we arrive at the following estimate for the second cluster integral:

$$-\beta_2 = (d_1 + d_2)(d_2 + d_3)(d_3 + d_1)\{O(l^2 d) + l_1 l_2 l_3 (\sin \gamma_{12} / \sin \phi_4)\};$$

$$(\phi_4 > d/l). \quad (36)$$

By the theorem of sine proportions, valid for spherical triangles, the quotient of the two sines is a symmetrical function of the three directions. The angle ϕ_4 vanishes (or equals π) whenever the three directions (a_1, a_2, a_3) are coplanar, whereby a_1 and a_2 are normally not parallel. For such directions, the estimate (36) becomes infinite; as we have mentioned above, the estimate (35) is then valid instead.

In computing the average of β_2 over all combinations of directions (a_1, a_2, a_3) we find that the combinations of directions which are coplanar within an angle $\pm\phi_m$ do form a fraction of the order ϕ_m of the total. Thus,

when we integrate EQUATION 36 over all other orientations the average of the sine ratio will be of the order

$$-\log \phi_m + \text{const.}$$

When we put $\phi_m = d/l$ and substitute the estimate (35) for coplanar orientation, the added term does not change the order of magnitude, and we still obtain

$$-\beta_2 = O(d^2 l^3 (\log (l/d) + \text{const})). \quad (37)$$

No concerted effort has been made to render this estimate more definite. As it is, the result

$$\beta_2/(\beta_1)^2 = O((d/l) \log (l/d)) \quad (38)$$

will offer some justification for the procedure which we shall, perforce, adopt in the following, where order corrections which depend on β_2 and higher cluster integrals will be neglected altogether. If we talk about "concentrated" solutions whenever $\beta_1 N/V$ is of the order unity or greater, then we may hope that our results will describe fairly concentrated *isotropic* solutions of rod-shaped particles reasonably well. The results for anisotropic solutions will be somewhat doubtful in all cases, and more so the more concentrated the solutions. Where plate-like particles are concerned, our approximations will introduce more serious errors, and we can hardly hope for more than that our result will describe concentrated solutions of such particles qualitatively rather than quantitatively.

We shall inquire, next, about the effects due to the finite range of the electrostatic repulsion between the particles.

We have mentioned before that the force (per unit area) between two parallel plates varies exponentially with the distance, and that the law of force for a different geometry is only modified by the effects due to divergence of the electric force-lines. The data needed for an exact prediction of the forces are not available, and even if we had them it would be a difficult and laborious task to compute the forces. But fortunately the resulting uncertainty will not, as a rule, count for much in the computation of the cluster integrals. Only the cases where very few cations or very few anions are present (or very few of either sign) might well require careful separate analysis. The most important modification of our previous results (EQUATIONS 30 and 31 a-h) will occur for long rods (EQUATION 31 h), in which case the effect of the electrostatic repulsion will be equivalent to an increase of the effective *diameter*. A similar increase of the effective *length* will cause a relatively insignificant increase of the covolume, unless the concentration of electrolyte is so low that κl is of the order unity, in which case, the problem of the repulsive forces must be reconsidered as a whole.

For the law of force between two cylindrical particles of the same diameter $d_1 = d_2 = d$, whose mantles are separated by a distance x , we now assume

$$w/kT = A(\gamma)e^{-\kappa x} \quad (39)$$

(of EQUATION 14 and pertinent discussion). More precisely, we assume that w has the value given by EQUATION 39 whenever the two cylinders cross, in the sense that the projections of their axes upon the plane parallel to both intersect. Such configurations yield the leading term

$$(d_1 + d_2)l_1l_2 \sin \gamma = 2dl_1l_2 \sin \gamma$$

of the excluded volume for long cylinders (the central parallelopiped of the solid figure illustrated in FIGURE 7, and compare EQUATION 30). The first cluster integral β_1 is defined by EQUATIONS 18 and 20. Assuming $w = w(x)$ as given by EQUATION 39 for all "crossed" configurations, the part of β_1 due to such configurations is simply

$$\beta_1(\text{crossing}) = 2l_1l_2 \left\{ -d + \int_0^\infty (e^{-w(x)/kT} - 1) dx \right\} \sin \gamma. \quad (40)$$

The consequent correction to the effective diameter, assuming EQUATION 39, is accordingly

$$\begin{aligned} \int_0^\infty (1 - e^{-Ae^{-\kappa x}}) dx &= \int_0^A (1 - e^{-u}) du/\kappa u \\ &= \kappa^{-1} \left(\log A + C + \int_A^\infty e^{-u} du/u \right). \end{aligned} \quad (41)$$

Here C denotes Euler's constant

$$\int_0^1 (1 - e^{-u}) \frac{du}{u} - \int_1^\infty e^{-u} \frac{du}{u} = C = -\Gamma'(1) = 0.5772 \dots \quad (41a)$$

For reasonably large values of A the exponential integral in EQUATION 41 may be neglected, and we get simply

$$\begin{aligned} -\beta_1(\text{crossing}) &= 2l_1l_2 d_{\text{eff}}(\gamma) \sin \gamma \\ &= 2l_1l_2 \{d + \kappa^{-1}(C + \log A(\gamma))\} \sin \gamma. \end{aligned} \quad (42)$$

Thus, the effective diameter equals the actual diameter increased by the distance δ at which the condition

$$w(\delta) = kT e^{-\sigma} = 0.561 kT \quad (43)$$

is satisfied.

In deriving this important rule, we have made certain physical assumptions and mathematical approximations; but, as long as κd is reasonably small (less than unity), the errors thus incurred ought to be very modest and EQUATION 42 should describe a good estimate. Moreover, the result is rather insensitive to modifications of the geometry, so that the required modifications of the first four terms of EQUATION 30 may be estimated in a

similar manner. We shall be content to observe that the increments of the effective lengths and diameters are of the same order of magnitude for all terms. This observation together with EQUATIONS A 14 and A 15 should be helpful in case an estimate of the end-corrections for long rods (which we shall neglect in the following) should be desired.

In EQUATION 42 we have indicated that the force constant A is expected to vary with the angle of intersection γ . A precise specification of that variation is contained in our EQUATION 13, and, in spite of the several approximations involved in the derivation, the relation

$$A(\gamma) = A(\pi/2)/\sin \gamma \quad (44)$$

ought to be very nearly true, with the one exception that, for small angles $\gamma < d/l_{12}$ the factor $1/\sin \gamma$ must be replaced by a smaller number of the order l_{12}/d , where l_{12} denotes the length of the overlap between the two cylinders.

If we disregard the exception just mentioned and neglect terms which represent end-wise approach of the particles, (corresponding to the first four terms in EQUATION 30), we may write EQUATION 42 in the form

$$\begin{aligned} -\beta_1(\gamma) = 2l_1l_2\{[d + \kappa^{-1}C + \kappa^{-1}\log A(\pi/2)] \sin \gamma \\ - \kappa^{-1}(\sin \gamma) \log (\sin \gamma)\}. \end{aligned} \quad (45)$$

The average of the excluded volume over all orientations equals

$$\begin{aligned} -\bar{\beta}_1 = 2\bar{b}_{12} = -\int \beta_1(\gamma) d\Omega/4\pi = -\frac{1}{\pi} \int_0^\pi \beta_1(\gamma) \sin \gamma d\gamma \\ = (\pi/2)l_1l_2(d + \bar{\delta}), \end{aligned} \quad (46)$$

with

$$\bar{\delta} = \kappa^{-1}[C + \log A(\pi/2) + \log 2 - \frac{1}{2}] = \kappa^{-1}[0.7704 + \log A(\pi/2)]. \quad (47)$$

The integral

$$\int_0^\pi \log (\sin \gamma) \sin^2 \gamma d\gamma,$$

which enters into the computation, might seem difficult, but it is easily computed from the Fourier series

$$-\log |2 \sin \gamma| = \cos 2\gamma + \frac{1}{3} \cos 4\gamma + \frac{1}{5} \cos 6\gamma + \cdots.$$

Concerning the absolute value of the force-constant A , we refer back to EQUATION 14 with attendant discussion and references.

Our results (33) and (46) for straight rods should apply without change to bent rods and *flexible chains* as long as they are not so tightly coiled that multiple contacts between pairs of different chains will be common. The statistics of such multiple contacts has not been investigated. In addition,

it stands to reason that, when the particles are so slender as to be very flexible, the effective range of the electrostatic repulsion will constitute the main part of their effective diameters.

We shall deal summarily with the case of thin, plate-shaped particles. According to EQUATION 30, the mutual excluded volume for a pair of such particles is practically independent of their thickness, barring only nearly parallel orientations of the particles. Then, if only

$$\kappa d \gg 1, \quad (48)$$

which condition excludes very low concentrations of electrolyte at the most, an increase of the effective diameter by a distance δ , determined according to EQUATION 43, will make very little difference. For the case of very low electrolyte concentration, the question of the forces would seem to require a more careful analysis than we have available at present.

Thermodynamic Properties of Isotropic Solutions. We shall be generally content with the first order corrections to the laws of ideal solutions. Accordingly, we abbreviate the expansion (21) as follows

$$\begin{aligned} \log B_p &= N_p(1 + \log(V/N_p) + \tfrac{1}{2}\beta_1(N_p/V)) \\ &= N_p(1 + \log(V/N_p) - b(N_p/V)), \end{aligned} \quad (48)$$

whereby, for a monodisperse solution of rod-shaped particles of length l and diameter d according to EQUATION 46

$$b = (\pi/4)l^2(d + \delta), \quad (49)$$

including a correction δ for the "padding" due to an ionic double layer. For a solution of plate-shaped particles EQUATION 48 is also valid over a more limited range of concentrations with the different value

$$b = (\pi/4)^2 d^3 \quad (50)$$

for the covolume, as given by EQUATION 34 when specialized to the case $l = 0$.

The variation of the free energy with the particle concentration can be obtained by substituting the result (48) in EQUATION 5. We are particularly interested in the derived quantities. We obtain from EQUATIONS 6 and 7, respectively,

$$P = kT\{(N_p/V) + b(N_p/V)^2\} \quad (51)$$

for the osmotic pressure and

$$\mu_p = \mu_p^0 + kT\{\log(N_p/V) + 2b(N_p/V)\} \quad (52)$$

for the chemical potential. These correspond to well-known formulas in the theory of gases; the salient point of the present theory is that the covolumes may be much greater than the actual volumes of the particles. We

also get reasonably simple results for a polydisperse solution which contains rod-shaped particles of various lengths l_1, \dots, l_s, \dots , but of identical diameters $d_1 = \dots = d_s = \dots = d$, and otherwise sufficiently similar, so that the effective diameter for any pair of particles is always $d + \bar{\delta}$. For this purpose, we substitute the covolumes given by EQUATION 46 in the more general formula (22), which yields

$$\begin{aligned} \log B_p &= \sum_s N_s (1 + \log (V/N_s)) - (\pi(d + \bar{\delta})/4V) \sum_{s,s'} N_s N_{s'} l_s l_{s'} \\ &= \sum_s N_s (1 + \log (V/N_s)) - (\pi/4)(d + \bar{\delta})(L^2/V), \end{aligned} \quad (53)$$

for the configuration integral, with the abbreviation

$$L = \sum_s N_s l_s \quad (54)$$

for the sum of the lengths of all particles present. The osmotic pressure is accordingly

$$P = kT \left\{ \sum_s (N_s/V) + (\pi/4)(d + \bar{\delta})(L/V)^2 \right\}, \quad (55)$$

and we get

$$\mu_s = \mu_s^0 + kT \log (N_s/V) + 2kT(\pi/4)(d + \bar{\delta})(L/V)l_s \quad (56)$$

for the chemical potential of the particles of length l_s .

The corresponding formulas for a polydisperse solution of circular plate-shaped particles are almost equally simple. We specialize EQUATION 32 to the case $l_1 = l_2 = 0$:

$$-\bar{\beta}_1(0, d_1; 0, d_2) = 2b = (\pi/4)^2 d_1 d_2 (d_1 + d_2);$$

then, with the abbreviations

$$\begin{aligned} D &= \sum_s N_s d_s \\ A &= (\pi/4) \sum_s N_s d_s^2 \end{aligned} \quad (57)$$

for the sums of all the diameters viz. areas of all particles present, we obtain

$$\log B_p = \sum_s N_s (1 + \log (V/N_s)) - (\pi/4)(DA/V) \quad (58)$$

$$P = kT \left\{ \sum_s (N_s/V) + (\pi/4)(DA/V^2) \right\} \quad (59)$$

$$\mu_s = \mu_s^0 + kT \log (N_s/V) + kT[A d_s + D(\pi/4) d_s^2](\pi/4V). \quad (60)$$

These results for plates might well have qualitative rather than quantitative significance. While certain colloids (bentonite) are known to consist of sheet-like particles, it is not known whether the outlines of the sheets are regular curves or polygons that might be reasonably approximated by circular disks. Nevertheless, it seems worth pointing out, that for a given total

area of the particles, both terms in the formula (59) for the osmotic pressure do increase as the degree of dispersion increases. (When all particles are cut into quarters the sum of the diameters is doubled and the total number of particles is increased by a factor of four.)

Returning to our result (55) for the osmotic pressure of *rod-shaped* particles, we note that the absolute value of the second term, which represents the deviation from the value appropriate to ideal solutions, is quite *independent of the subdivision into individual lengths*. (On the other hand, a lengthwise splitting of the particles, if possible, will increase both terms in EQUATION 55.) On this basis, we should be prepared to find that the osmotic pressures of *flexible* chain-like particles in concentrated solutions may be practically independent of the subdivision of the chains. For rigid rod-shaped particles, we do not anticipate this phenomenon (in the isotropic phase, anyway), because, as we shall show next, such solutions will form an anisotropic phase as soon as the ratio (total covolume/volume) exceeds a certain critical value.

Anisotropic Solutions. We shall investigate the possibility that a solution of rod-shaped particles may form a nematic *liquid crystal* in which the distribution of orientations of the particles is anisotropic, while the distribution of the particles in space is homogeneous, and does not exhibit the periodic variation of density which characterizes solid crystals (periodicity in three dimensions) and *smectic* liquid crystals (periodicity in one dimension). We shall show that the concentration of particles need not be so very large (in terms of actual volume occupied) before the isotropic solution becomes unstable, relative to an anisotropic phase of the nematic type. Whether the latter will be stable, relative to other types of anisotropic phases, is a question which involves much more difficult computations, and we shall not try to settle it.

We introduce a distribution-function $f(\mathbf{a})$ for the directions \mathbf{a} of the axes of the cylindrical particles, normalized according to EQUATION 26. When we neglect the terms which depend on β_2 and higher cluster integrals in the expansion given by EQUATION 27, we arrive at the following formula for the configuration-integral

$$\log B_p = N_p \{1 + \log (V/N_p)\} - \int f(\mathbf{a}) \log 4\pi f(\mathbf{a}) d\Omega(\mathbf{a}) \quad (61)$$

$$+ (N_p/2V) \int \beta_1 (\cos^{-1} (\mathbf{a} \cdot \mathbf{a}')) f(\mathbf{a}) f(\mathbf{a}') d\Omega d\Omega'.$$

The function $f(\mathbf{a})$ is implicitly determined by the condition

$$B_p = \text{maximum}, \quad (62)$$

(subject to the restriction (26)).

We shall introduce convenient abbreviations for the two functionals which enter into EQUATION 61:

$$\sigma(f) = \int f(\mathbf{a}) \log 4\pi f(\mathbf{a}) d\Omega(\mathbf{a}), \quad (63)$$

$$-2b\rho(f) = \tilde{\beta}_1 \rho(f) = \int \beta_1(\cos^{-1}(\mathbf{a} \cdot \mathbf{a}')) f(\mathbf{a}) f(\mathbf{a}') d\Omega d\Omega'; \quad (64)$$

where, in conformity with EQUATION 46, we understand:

$$-2b = \tilde{\beta}_1 = \int_0^{\pi/2} \beta_1(\gamma) \sin \gamma d\gamma. \quad (64a)$$

In addition, we shall denote the concentration of particles by

$$c = (N_p/V). \quad (65)$$

In this shorthand, the condition (62) becomes

$$\sigma(f) + bc\rho(f) = \text{minimum}, \quad (66)$$

wherein f is subject to the restriction

$$\int f d\Omega = 1. \quad (26)$$

The value of the minimum required by the condition (66) determines the free energy of the system according to EQUATIONS 5 and 61:

$$\begin{aligned} F(\text{solution}) - F(\text{solvent}) &= N_p \mu_p^0 - kT \log B_p \\ &= N_p \mu_p^0 + N_p kT \{ \log c - 1 + \sigma(f) + bc\rho(f) \}. \end{aligned} \quad (67)$$

We may apply Lagrange's method to the problem (66), thus

$$\delta\sigma(f) + bc\delta\rho(f) - \lambda\delta \int f d\Omega = 0. \quad (68)$$

The usual manipulations lead to the non-linear integral equation

$$\log(4\pi f(\mathbf{a})) = \lambda - 1 + c \int \beta_1(\mathbf{a}, \mathbf{a}') f(\mathbf{a}') d\Omega'. \quad (69)$$

EQUATION 69 is satisfied by every function which renders the functional of the problem (66) stationary; the true solution of (66) is included among these. The constant function

$$f = f_0 = 1/4\pi, \quad (70a)$$

which describes the isotropic distribution, is always a solution of EQUATION 69, with

$$\sigma(f_0) = 0; \quad \rho(f_0) = 1; \quad \lambda = 1 + 2bc \quad (70b)$$

On the other hand, in order to show that for sufficiently large values of c the solution (70) will not be the true solution of the problem (66), we only

have to find some function f_1 such that $\sigma(f_1)$ is finite and $\rho(f_1) < 1$; then, when we take c large enough, the inequality

$$\sigma(f_1) + bc\rho(f_1) < \sigma(f_0) + bc\rho(f_0) = bc$$

can certainly be satisfied.

According to our previous considerations, the function $-\beta_1(\gamma)$ is an increasing function of $(\sin \gamma)$, so that a trial function f_1 with the required properties can be constructed very simply as follows: We choose an angle γ_1 such that

$$-\beta_1(\gamma) < -\tilde{\beta}_1; \quad \gamma < 2\gamma_1,$$

a preferred direction \mathbf{a}_0 and the following trial function

$$f_1(\mathbf{a}) = 0; \quad |(\mathbf{a}_0 \cdot \mathbf{a})| < \cos \gamma_1$$

$$f_1(a) = 1/4\pi(1 - \cos \gamma_1); \quad \cos \gamma_1 < |(\mathbf{a}_0 \cdot \mathbf{a})| < 1.$$

Some of the unwanted solutions of EQUATION 69—possibly all of them—may be interpreted as solutions of a modified variation problem:

$$\begin{aligned} \int f d\Omega &= 1 && \text{prescribed} \\ \sigma(f) &= \sigma_1 \geq 0 \\ \rho(f) &= \text{minimum} = \rho_m(\sigma_1). \end{aligned} \quad (71)$$

This leads again to EQUATION 68 with the difference that c is interpreted as a Lagrange multiplier on par with λ .

The second restriction in the problem (71) is in effect no different than the inequality

$$\sigma(f) \geq \sigma_1, \quad (71a)$$

because the function $\beta(\gamma)$ in EQUATION 64 is continuous (less would suffice). In consequence, if we know one function $f(\mathbf{a})$, which realizes a certain value of ρ , we can always find another which realizes very nearly the same value of ρ , but gives us a *greater* value of σ . All we have to do is introduce a very rapid local fluctuation of $f(\mathbf{a})$. This reasoning leads to the inequality

$$(\rho_m(\sigma_2) - \rho_m(\sigma_1))/(\sigma_2 - \sigma_1) \leq 0; \quad (71b)$$

in words: the minimum of ρ is a never-increasing function of σ .

One way to solve the problem (66), at least in principle, is to solve the more general problem (71) first for all values of σ . For greater flexibility, we may describe the resulting relation between ρ_m and σ in parameter form

$$\sigma = \sigma(\alpha); \quad \rho_m = \rho(\alpha); \quad (72a)$$

then the solution of EQUATION 66 must satisfy the condition

$$\sigma'(\alpha) + bc\rho'(\alpha) = 0, \quad (72b)$$

and, in order that the state thus described be stable

$$\sigma''(\alpha) + bc\rho''(\alpha) > 0. \quad (72c)$$

According to these results, if the function

$$(-d\rho_m/d\sigma) = -\rho'_m(\sigma) \quad (73)$$

is a steadily decreasing function of σ , then the transition from the isotropic to the anisotropic phase will be continuous and take place at the concentration given by the condition

$$1 + bc(d\rho_m/d\sigma)_{\sigma=0} = 0. \quad (74)$$

On the other hand, if the function (73) increases for small values of σ , reaches a maximum (as it must because $\rho > 0$), and decreases thereafter, then the isotropic solution will become unstable towards finite disturbances at some concentration lower than that required by EQUATION 74. In this case, the anisotropic phase will always possess a finite degree of anisotropy. Moreover, there will be a pair of concentrations for which the two phases can coexist; a solution of concentration intermediate between these will separate into two phases.

It is possible to show by rather general qualitative reasoning that the second alternative—a discontinuous transition—must be realized when an anisotropic solution is formed. We may as well assume that the anisotropic phase has cylindrical symmetry around some preferred direction \mathbf{a}_0 ; this restriction is unimportant, because it allows the distribution-function to contain spherical harmonics of all (even) orders. The odd orders are excluded if we assume that the solution is not polar in the crystallographic sense (seignette-electric):

$$f(\mathbf{a}) = f(-\mathbf{a}).$$

Under these assumptions, $f(\mathbf{a})$ may be developed in a series of even Legendre polynomials

$$4\pi f(\mathbf{a}) = 1 + 5A_2P_2(\mathbf{a} \cdot \mathbf{a}_0) + 9A_4P_4(\mathbf{a} \cdot \mathbf{a}_0) + \dots \quad (75)$$

Moreover, since the homogeneous quadratic functional defined by EQUATION 64 is invariant against all rotations of the frame of reference, the Legendre polynomials are its eigenfunctions and an expansion of the type

$$\rho(f) = 1 - B_2A_2^2 - B_4A_4^2 - \dots \quad (76)$$

is valid. Only even powers of A_2, A_4, \dots , (in fact, only the second powers), occur in EQUATION 76. However, when we substitute the expansion (75) in EQUATION 73, we get

$$\begin{aligned} \sigma(f) &= (5/2)A_2^2 + (9/2)A_4^2 + \dots \\ &\quad - (25/21)A_2^2 - (45/7)A_2^2A_4 - \dots \\ &\quad + (125/28)A_2^4 + \dots \end{aligned} \quad (77)$$

Linear terms do not occur in this expansion either, but *every cubic term is present*. The critical condition defined by EQUATION 74 is fulfilled by the smallest value (c_0) of c , which causes any one of the coefficients in the expansion

$$((5/2) - B_2bc)A_2^3 + ((9/2) - B_4bc)A_4^3 + \dots$$

to vanish. It does not matter which one vanishes first; for the sake of argument let us assume that it is the coefficient of A_2^3 . Then, if we take

$$f(\mathbf{a}) = 1 + A_2 P_2(\mathbf{a} \cdot \mathbf{a}_0)$$

we have the expansion (convergent for $|A_2| < 1$):

$$\sigma(f) + bc_0\rho(f) = bc_0 - (25/21)A_2^3 + (125/28)A_2^4 + \dots$$

For finite, not too large, positive values of A_2 , the sum of this expansion certainly takes values smaller than bc_0 . This means that the isotropic solution becomes unstable towards finite disturbances at some concentration lower than that required by EQUATION 74 for a continuous transition, so that a *discontinuous transition* must take place at some lower concentration.

We shall try next, to get some idea about the distribution of orientations of the particles in the anisotropic solutions, and to estimate the thermodynamic functions for these solutions. The variation problem (66) is best attacked directly: the technique is to construct plausible trial functions with as many variable parameters as one can handle conveniently; the parameters are then adjusted so as to approach the required minimum as closely as possible.

As regards the general nature of the distribution, we note that $-\beta_1(\mathbf{a}, \mathbf{a}')$ has a minimum when the directions (\mathbf{a} , \mathbf{a}') are parallel and a maximum when they are perpendicular. The smallest possible value of $\rho(f)$ is, therefore, attained when all particles have exactly the same orientation; according to our somewhat approximate formula (45), the function $\rho(f)$ then vanishes. For this singular distribution, however, $\sigma(f)$ becomes infinite. As a compromise, we have to expect a distribution which is more or less concentrated around a preferred direction \mathbf{a}_0 , with this direction as an axis of symmetry. EQUATION 69 indicates, in a general way, how the density will decrease with the angle

$$\Theta = \cos^{-1}(\mathbf{a} \cdot \mathbf{a}_0) \quad ; \quad (78)$$

for large angles an exponential function of $(\sin \Theta)$ is thus indicated. A trial function of the type

$$\text{constant} \times (\cosh(\alpha \sin \Theta))^{-n} \quad (79)$$

would therefore seem promising, and tentative computations for $n = 3$ gave encouraging results, but this lead was abandoned on account of the effort involved. The simpler function

$$f(\mathbf{a}) = (\alpha/4\pi \sinh \alpha) \cosh(\alpha(\mathbf{a} \cdot \mathbf{a}_0)) = (\alpha/4\pi \sinh \alpha) \cosh(\alpha \cos \Theta) \quad (80)$$

decreases rather too rapidly for large angles, and it contains but one parameter. It was, nevertheless, adopted as the best tractable function. Even so, according to EQUATION 67, the function to be minimized for the problem (66) is the *free energy* itself, so that our results for this important thermodynamic function ought not to be very much in error.

Further, to simplify the computations, we shall allow the approximate description

$$-\beta_1(\gamma) \sim 2l_1 l_2 (d + \delta) \sin \gamma = (8/\pi) b \sin \gamma \quad (81)$$

of the more complicated function given by EQUATION 45. When EQUATIONS 80 and 81 are substituted in EQUATION 64, the resulting integral can be evaluated in terms of elementary functions, together with a Bessel function of order 2, and imaginary argument. For the definition (B 18) and properties of this function, and for the details of the integration, we refer to the Appendix, Section 2. We quote here the result for a monodisperse solution

$$\rho(\alpha) = 2(\sinh \alpha)^{-2} I_2(2\alpha) \quad (82)$$

(cf. EQUATION B 17). The evaluation of the integral in EQUATION 63 for the function (80) is elementary and yields

$$\sigma(\alpha) = \log(\alpha \coth \alpha) - 1 + (\sinh \alpha)^{-1} \tan^{-1}(\sinh \alpha). \quad (83)$$

The power series

$$4\rho(\alpha) = 4 - (\alpha^4/90) + (2\alpha^6/945) - (71\alpha^8/226800) + \dots; \quad (|\alpha| < \pi) \quad (84)$$

$$\sigma(\alpha) = (\alpha^4/90) - (2\alpha^6/810) + (108\alpha^8/226800) - \dots; \quad (|\alpha| < \pi/2),$$

illustrate our consideration of continuous vs. discontinuous transition (when an expansion of the type (75) is constructed for the function (80), the coefficient of P_2 is of the order (α^2)). However, the values of α which correspond to stable anisotropic solutions are far outside the limits of convergence of the series (84). Fortunately, the required values of α are so large that the asymptotic representations

$$\sigma(\alpha) \sim \log \alpha - 1, \quad (85)$$

$$\rho(\alpha) \sim 4(\pi\alpha)^{-1/2} \{1 - 30(32\alpha)^{-1} + 210(32\alpha)^{-2} + 1260(32\alpha)^{-3} + \dots\}, \quad (86)$$

(cf. B 20), are eminently suitable for computation. The condition (72b) for internal equilibrium becomes, after rearrangement,

$$(\pi\alpha)^{1/2} = 2bc \{1 - 90(32\alpha)^{-1} + 1050(32\alpha)^{-2} + 8820(32\alpha)^{-3} + \dots\}. \quad (87)$$

The asymptotic behavior of α for high concentrations is evident from the inverted series

$$\alpha \sim (4/\pi) (bc)^2 - (45/8) + O((bc)^{-2}). \quad (88)$$

As a measure of the spread in angle, we may compute the mean square of $\sin(\Theta/2)$, if we understand by Θ , (this time), the angle between the direction \mathbf{a} of a particle and the nearer of the two directions $(+\mathbf{a}_0, -\mathbf{a}_0)$. With this convention, we have

$$(2 \sin (\Theta/2))^2 = (2/\alpha) \coth \alpha \sim 1/\alpha. \quad (89)$$

The standard deviation of the angle Θ is, therefore, about proportional to $\alpha^{-1/2}$ or, if we disregard the higher terms in EQUATION 87, inversely proportional to the concentration:

$$(2 \sin (\Theta/2))^2 \sim (\pi/4) (bc)^{-2} \quad (89a)$$

Combination of EQUATIONS 85, 86, and 87 yields for high concentrations

$$bc\rho \sim 2 + 75\pi(8bc)^{-2} + \dots \quad (90)$$

$$\sigma \sim \log(4/\pi) - 1 + 2 \log(bc) - 90\pi(8bc)^{-2} + \dots \quad (91)$$

The free energy is given by EQUATION 67; in view of EQUATION 72b we derive therefrom the following simple *general* formula for the *osmotic pressure*

$$P = -(\partial F/\partial V)_{N,p} = kTc(1 + bc\rho). \quad (92)$$

In particular, if we take EQUATION 90 seriously

$$P = kT c(3 + 75\pi(8bc)^{-2} + \dots). \quad (93)$$

On this basis, the osmotic pressure should be just a little greater than three times the ideal pressure.

The simple results (89a) and (93), however, depend on a rather severe over-simplification of the physical picture. The approximation (81) for EQUATION 45 tends to overestimate the deviation of the angles for high concentrations. The several approximations and simplifications which enter into EQUATION 93 would tend to make this an underestimate, possibly a very bad one; it is barely conceivable that our neglect of the attractive van der Waal's (dispersion) forces might in some cases bring about a measure of compensation.

For the lowest concentrations at which the anisotropic solutions can exist, the approximations which lead to EQUATIONS 85 and 86 may still be quite tolerable. In those cases it is best to solve EQUATION 87 numerically, because the expansion is only semi-convergent and the inversion (88) aggravates its tendency to diverge. In constructing tables or graphs there is, of course, no need for the inversion; it is just as well to compute σ , ρ and c all as functions of the parameter α .

We still have to compute the conditions for equilibrium of the two liquid phases. The osmotic pressure of the anisotropic solution is given by EQUATION 92. The chemical potential is computed according to the definition (7) by differentiation of EQUATION 67, whence in view of EQUATIONS 65 and 72b:

$$\mu_p = (\partial F / \partial N_p)_T = \mu_p^0 + kT \{ \log c + \sigma + 2bc_p \}. \quad (94)$$

The equilibrium conditions (8) require that the functions given by (92) and (94) be equated with the corresponding functions (51) and (52) for the isotropic phase. The two concentrations c_a and c_i are thus determined by the two equations

$$\begin{aligned} c_a + bc_a^2 &= c_i + bc_i^2 \\ \log c_a + \sigma + 2bc_a\rho &= \log c_i + 2bc_i, \end{aligned}$$

which we may write in terms of the total covolumes as follows:

$$\begin{aligned} bc_a(1 + bc_a\rho) &= bc_i(1 + bc_i) \\ \log(bc_a) + \sigma + 2bc_a\rho &= \log(bc_i) + 2bc_i. \end{aligned} \quad (95)$$

Here, σ and ρ are functions of (bc_a) described implicitly by EQUATIONS 85, 86, and 87. The following results were obtained by numerical solution of the system of equations:

$$\begin{aligned} \alpha &= 18.584 & bc_i &= 3.3399 \\ \rho &= 0.49740 & bc_a &= 4.4858 \\ \sigma &= 1.9223 & \rho bc_a &= 2.2313 \\ c_a/c_i &= 1.343. \end{aligned} \quad (96)$$

The standard deviation of $\sin(\Theta/2)$ given by EQUATION 89 corresponds to an angle $\Theta = 13.3^\circ$.

It is a matter of interest to see how the expansions (88), (90), and (91) work out in the worst possible case. The values obtained from these formulae, as abbreviated, for the case $bc_a = 4.486$,

$$\alpha = 19.9; \quad \rho bc_a = 2.184; \quad \sigma = 2.021,$$

may be compared with the accurately computed values (96).

It would take us too far to develop a theory for the anisotropic phase of a polydisperse solution. The difficulty is that long rods will be more perfectly oriented than short rods, so that one has to compute a whole set of mutually dependent distribution-functions, one for each size of particles. Moreover, each composition of the anisotropic phase presents a separate problem of this type. Nevertheless, the mathematical analysis in the Appendix has been kept as general as was feasible, in order to facilitate computations for polydisperse systems.

It is possible to foresee that when two phases are formed, the longest particles will collect preferentially in the anisotropic phase, and that the total concentrations in each phase will vary with the ratio of the volumes.

Possibly, the best experimental tests of the present theory will consist in measurements of light scattering. As is well known, the light scattering per particle is inversely proportional to (dP/dc) . However, when the longest dimensions of the particles are comparable to that of the measuring light, this simple relation applies only to *scattering at small angles*. We have, in effect, shown that the presence of one particle reduces the density of scattering matter up to a distance which equals the length of a presumptive neighbor, so that the *phase relations* of the light waves scattered from *different particles* must be considered in the interpretation of the large angle scattering.

On the basis of our results (92) and (51), the *anisotropic solution* ought to scatter more light at small angles than the isotropic phase in equilibrium with it. Predictions for large angles must await a mathematical analysis of the optical problems; the distribution of scattering matter around any one particle obviously depends on the degree of orientation of the particles.

Appendix

The Mutual Excluded Volume of Two Cylinders

We first compute the excluded volume $-\beta_1(0, d_1; 0, d_2)$ for two circular plates of vanishing thickness and diameters $d_1 = 2r_1$; $d_2 = 2r_2$. Besides, this is one of the interesting limiting cases.

Let the first plate be fixed with its center at the origin and let its normal form an angle γ with the y axis, in the (y, z) plane. We allow the second plate to move, but we keep its orientation constant, so that its normal is always parallel to the y axis. When we require that the two plates must not intersect, what is the volume inaccessible to the center of the second plate?

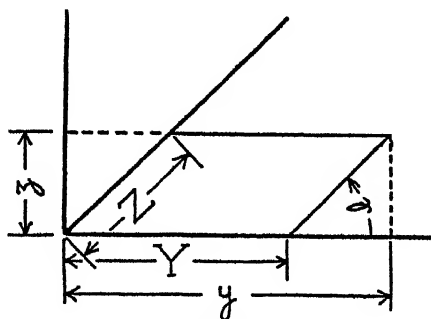


FIGURE 1.

Some of the analysis becomes a little simpler if we replace the coordinates (y, z) by the pair (Y, Z) referred to skew axes parallel to the plates 2, 1 respectively:

$$\begin{aligned} y &= Y + Z \cos \gamma \\ z &= Z \sin \gamma \end{aligned} \quad (A 1)$$

Then the intersection of the excluded region with the plane

$$z = Z \sin \gamma = \text{const.}$$

is bounded by the curve formed by the center of circle 2 as this circle rolls on a chord AB of the circle 1 (see FIGURE 2).

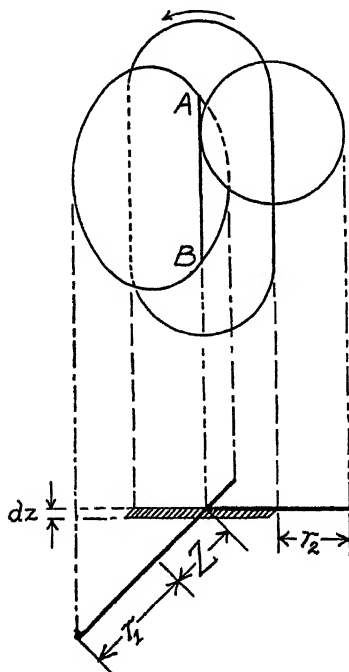


FIGURE 2.

The length of the chord equals

$$2(r_1^2 - Z^2)^{1/2}.$$

The area of the intersection is, evidently,

$$A(Z) = 4(r_1^2 - Z^2)^{1/2} r_2 + \pi r_2^2$$

(rectangle + two half circles), and the excluded volume is, accordingly,

$$\begin{aligned} -\beta_1(0, d_1; 0, d_2) &= \int A(Z) dz = \int A(Z) dZ \sin \gamma \\ &= 2\pi r_1 r_2 (r_1 + r_2) \sin \gamma \\ &= (\pi/4) d_1 d_2 (d_1 + d_2) \sin \gamma. \end{aligned} \quad (\text{A } 2)$$

The excluded region, illustrated in FIGURE 3, is bounded by four planes and by the fourth degree surface described by the equations

$$(r_1^2 - Z^2)^{1/2} + (r_2^2 - Y^2)^{1/2} = \pm x. \quad (\text{A } 3)$$

(Here and in the following we always take positive roots). This surface joins the planes

$$\begin{aligned} Z &= \pm r_1 \\ Y &= \pm r_2 \end{aligned} \quad (\text{A } 4)$$

along the circles

$$x^2 + Y^2 = r_2^2$$

$$x^2 + Z^2 = r_1^2,$$

respectively, in such a manner that the normal directions are continuous except at the points

$$x = 0, \quad Y = r_1, \quad Z = r_2$$

where two such circles touch one another.

A curve bounding the intersection of the excluded region with a plane $x = \text{const.}$ is described by EQUATION A 3 for $x = \text{const.}$ or it consists of segments of curves described thus alternating with segments of the straight lines described by EQUATION A 4, according to the value of x . The cross-section for $x = 0$ is simply the parallelogram given by EQUATION 4. The four possible cases are illustrated in FIGURE 3.

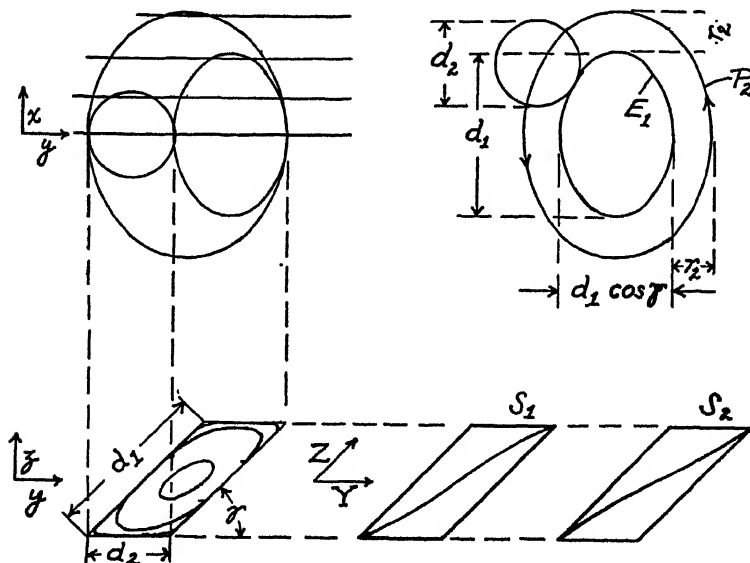


FIGURE 3.

On each curve of this type there is a pair of points for which

$$y = Y + Z \cos \gamma = \text{extremum } (S_2)$$

and another pair for which

$$z' = Y \cos \gamma + Z = \text{extremum } (S_1).$$

The loci of such points are space-curves S_1, S_2 on the surface of the excluded volume; their projections on the (y, z) plane are described by the equations

$$(Z/Y)(r_1^2 - Z^2)^{-1/2}(r_2^2 - Y^2)^{1/2} = \begin{cases} \cos \gamma; & (S_2) \\ 1/\cos \gamma; & (S_1) \end{cases} \quad (\text{A } 5)$$

The significance of S_2 will be clear from the observation that when the center of plate 2 is on S_2 , not only does this plate make rim-to-rim contact with plate 1, but (in addition), a cylinder raised perpendicularly on the rim of plate 2 also just touches the rim of plate 1. Similarly, when the center of plate 2 is on S_1 , then the rim of plate 2 just touches a cylinder raised perpendicularly on the rim of plate 1.

The curve S_2 separates the two parts of the surface of the excluded volume which are seen from opposite directions along the normal of plate (2). Its projection P_2 on the (x, y) plane (parallel to plate 2), which delimits the projection of the excluded volume, may be described as the locus of points whose distance from the nearest point of the projection of plate (1) is precisely r_2 . The projection of plate (1) on the plane of plate (2) is an ellipse E_1 of semi-axes $r_1, r_1 \cos \gamma$; if a circle of radius r_2 rolls on this ellipse its center traces P_2 .

The area $A(P_2)$ can be evaluated by a general method applicable to rolling-figures of continuous tangent. We let R denote the curve described by the center of a circle of radius r as the circle rolls on closed curve C . The algebraic sum of the curvatures of the circle and the curve C must be positive (convex) everywhere on C . The area between C and R is easily found as follows (FIGURE 4):

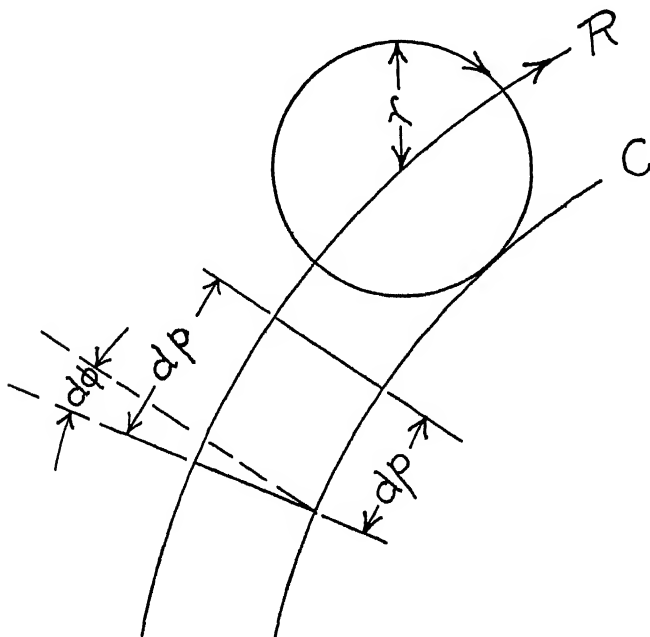


FIGURE 4.

Consider the area contained between two normals of directions $\varphi, \varphi + d\varphi$, the segments $dp(C)$ of the perimeter of C , and the segment

$$dp'(R) = dp(C) + r d\varphi$$

of the perimeter of R . The area of the trapezium thus formed equals

$$\frac{1}{2}r(dp + dp') = rdp + \frac{1}{2}r^2 d\varphi.$$

Integrating around C we find simply

$$A(R) - A(C) = \int r dp(C) + \frac{1}{2}r^2 \int d\varphi = rp(C) + \pi r^2 \quad (A.6)$$

where $p(C)$ denotes the circumference of C , and $A(C)$ its area. For the ellipse E_1 described above we have

$$A(E_1) = \pi r_1^2 |\cos \gamma| \quad (\text{A } 7)$$

$$p(E_1) = 4r_1 E(\sin \gamma) = 4r_1 \int_0^{\pi/2} (1 - \sin^2 \gamma \sin^2 \varphi)^{1/2} d\varphi$$

where the customary notation $E(\sin \gamma)$ is used for a complete elliptic integral of the second kind. We substitute these results in (A 6) along with $r = r_2$ and obtain for the projection P_1 :

$$A(P_2) = \pi r_1^2 |\cos \gamma| + 4r_1 r_2 E(\sin \gamma) + \pi r_2^2 \quad (\text{A } 8)$$

Similarly, of course,

$$A(P_1) = \pi r_2^2 + 4r_1 r_2 E(\sin \gamma) + \pi r_1^2 |\cos \gamma|. \quad (\text{A } 9)$$

Actually, the profile areas (A 8) and (A 9) are the only properties of S_1 and S_2 which will enter into our final result for the excluded volume of two cylinders, but some additional analysis has been included as an aid to visualization.

So far, we have considered the thin plates of diameters d_1, d_2 . Now, let us replace the second by a cylinder of length l_2 , diameter d_2 . Our solution for the case $l_2 = 0$ has an

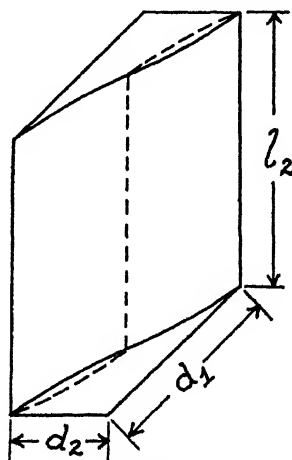


FIGURE 5.

important connection with the more general problem: As we consider the two halves of the surface illustrated in FIGURE 3 separated by the curve S_1 , the front half is the locus for the center of the rear end of cylinder (2) at closest approach to plate 1, and the rear half of the surface is the locus for the front end-center of the cylinder for the opposite type of end-wise contact. The corresponding loci for the center of cylinder (2) are displaced by distances $\pm \frac{1}{2} l_2$ in the direction of the cylinder axis. If the two end surfaces formed by these loci are joined rim to rim by a cylindrical mantle parallel to the axis of cylinder (2), then that mantle is the locus for the center of (2) when this cylinder makes lateral contact with the rim of plate (1).

The excluded volume for this case is the same as for $l_1 = l_2 = 0$, plus that of a cylinder of length l_2 and orthogonal section P_2 :

$$-\beta_1(0, d_1, l_2, d_2) = -\beta_1(0, d_1, 0, d_2) + l_2 A(P_2). \quad (\text{A } 10)$$

The end-faces of the inserted cylindrical piece are parallel, so that the computation of the volume is not affected by their complicated shape.

The final generalization to the case $l_1 > 0$ proceeds in a similar manner: The body illustrated by FIGURE 5 is cut as indicated by the dotted line. The cut, most simply taken perpendicularly to the plane of the projection, follows the separated halves of the curve S_1 and the median plane of the cylinder inserted in the previous step. A second cylindrical

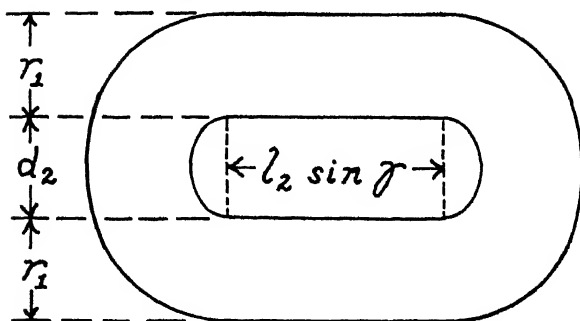


FIGURE 6

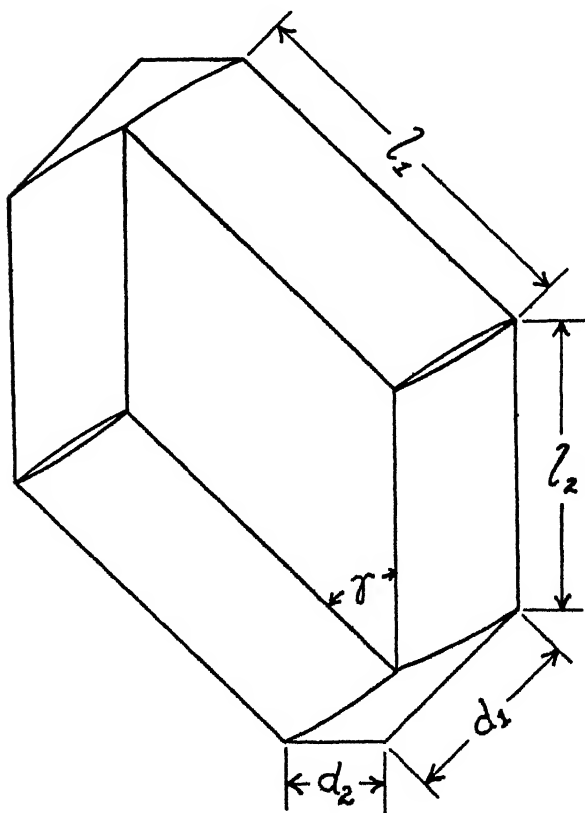


FIGURE 7

piece of length l_2 parallel to the normal of the plate (1) and cross-section, as indicated by FIGURE 6, is inserted. This time the added volume is

$$l_2 A(P_1) + 2 l_2 l_2 (r_1 + r_2) \sin \gamma,$$

and the volume of the resulting domain, illustrated in FIGURE 7, equals

$$\begin{aligned} -\beta_1(l_1, d_1; l_2, d_2; \gamma) &= (\pi/4)d_1 d_2(d_1 + d_2) \sin \gamma \\ &+ l_2[(\pi/4)d_2^2 + d_1 d_2 E(\sin \gamma) + (\pi/4)d_1^2 |\cos \gamma|] \\ &+ l_1[(\pi/4)d_1^2 + d_1 d_2 E(\sin \gamma) + (\pi/4)d_2^2 |\cos \gamma|] + l_1 l_2(d_1 + d_2) \sin \gamma. \end{aligned} \quad (\text{A } 11)$$

The integrations involved in the computation of the average

$$\int \beta_1(\gamma) d\Omega/4\pi = \frac{1}{2} \int_0^\pi \beta_1(\gamma) \sin \gamma d\gamma$$

should be obvious for all terms except those which involve the elliptic integral $E(\sin \gamma)$. These too, reduce to a very simple form by a suitable substitution. By the definition of the complete elliptic integral we have

$$4E(\sin \gamma) = 4 \int_0^{\pi/2} (1 - \sin^2 \gamma \sin^2 \psi)^{1/2} d\psi = \int_0^{2\pi} (1 - \sin^2 \gamma \sin^2 \psi)^{1/2} d\psi$$

Hence

$$4 \int_0^\pi E(\sin \gamma) \sin \gamma d\gamma = \int_0^\pi \sin \gamma d\gamma \int_0^{2\pi} d\psi (1 - \sin^2 \gamma \sin^2 \psi)^{1/2} \quad (\text{A } 12)$$

By the substitution

$$\sin \gamma \sin \varphi = \cos \vartheta$$

$$\cos \gamma = \sin \vartheta \sin \varphi,$$

(which may be interpreted as a change of polar coordinates in space), the integral (A 12) becomes

$$4 \int_0^\pi E(\sin \gamma) \sin \gamma d\gamma = \int_0^\pi \sin \vartheta d\vartheta \int_0^{2\pi} d\varphi \sin \vartheta = \pi^2 \quad (\text{A } 13)$$

With the aid of this result and obvious elementary integrations we find from EQUATION A 11:

$$\begin{aligned} - \int \beta_1(l_1, d_1; l_2, d_2; \gamma) d\Omega/4\pi &= (\pi/4)^2 d_1 d_2(d_1 + d_2) \\ &+ (\pi/4)(l_1 d_1^2 + l_2 d_2^2) + (\pi/8)(l_1 d_2^2 + l_2 d_1^2) \\ &+ (\pi^2/8)(l_1 + l_2) d_1 d_2 + (\pi/4) l_1 l_2 (d_1 + d_2) \end{aligned} \quad (\text{A } 14)$$

For a pair of cylinders of lengths l_1, l_2 , capped by hemispheres of diameters d_1, d_2 , the computation of the mutual excluded volume is quite analogous to the preceding. Several details are much simpler. For the case $l_1 = l_2 = 0$, the result (A 2) is replaced by the volume of a sphere of diameter $d_1 + d_2 = 2d$. The profiles (P_1, P_2) are replaced by circles of radius d (compare A 8, A 9). The assembled final result for capped cylinders is

$$-\beta_1(\gamma) = (4\pi/3)d^3 + \pi d^2(l_1 + l_2) + 2dl_1 l_2 \sin \gamma. \quad (\text{A } 15)$$

Here, the averaging over directions involves only the simple integral

$$\int \sin \gamma d\Omega/4\pi = \pi/4. \quad (\text{A } 16)$$

The Mean Covolume for Anisotropic Solutions

We shall show how the multiple integral

$$\bar{\beta}_{1\rho}(f_1, f_2) = \int f_1(a_1) f_2(a_2) \beta_1(\cos^{-1}(a_1 \cdot a_2)) d\Omega(a_1) d\Omega(a_2), \quad (\text{B } 1)$$

where \mathbf{a}_1 , \mathbf{a}_2 denote variable unit vectors which specify the orientations of particles, can be reduced to a single integral when the distribution-functions are of the special type

$$f(\mathbf{a}) = (\alpha_s/4\pi \sinh \alpha_s) \cosh \alpha_s(\mathbf{a} \cdot \mathbf{a}_0), \quad (\text{B } 2)$$

symmetrical about the discretion \mathbf{a}_0 (axis of the liquid crystal). Concerning β_1 we assume only

$$\beta_1(\gamma) = \beta_1(\pi - \gamma) = F(\sin \gamma) \quad (\text{B } 3)$$

at this stage, although in the end we shall introduce the approximation

$$\beta_1(\gamma) = \beta_1(\pi/2) \sin \gamma, \quad (\text{B } 4)$$

derived in the text.

Consider the integral

$$J = \int \cosh (\alpha_1(\mathbf{a}_1 \cdot \mathbf{a}_0) + \alpha_2(\mathbf{a}_2 \cdot \mathbf{a}_0)) F(\sin \gamma) d\Omega_1 d\Omega_2;$$

$$\cos \gamma = (\mathbf{a}_1 \cdot \mathbf{a}_2). \quad (\text{B } 5)$$

The value of this integral is not affected by the substitution

which changes the first factor of the integrand into

$$\cosh (\alpha_1(\mathbf{a}_1 \cdot \mathbf{a}_0) - \alpha_2(\mathbf{a}_2 \cdot \mathbf{a}_0)).$$

The arithmetic mean of the two integrals involves the factor

$$\cosh (\alpha_1(\mathbf{a}_1 \cdot \mathbf{a}_0)) \cosh (\alpha_2(\mathbf{a}_2 \cdot \mathbf{a}_0))$$

instead, and by comparison with EQUATIONS. B 1, B 2, we readily verify the identity

$$J = (4\pi)^2 (\sinh \alpha_1 \sinh \alpha_2 / \alpha_1 \alpha_2) \beta_{1\rho}(f_1, f_2). \quad (\text{B } 6)$$

We proceed to evaluate J .

For the direction \mathbf{a}_1 , we next introduce polar coordinates (Θ_1, ϕ_1) referred to \mathbf{a}_0 , but the direction \mathbf{a}_2 we shall specify in terms of polar coordinates (γ, ϕ) referred to the direction \mathbf{a}_1 , such that $\phi = 0$ for $\mathbf{a}_2 = \mathbf{a}_0$. Then

$$(\mathbf{a}_0 \cdot \mathbf{a}_1) = \cos \Theta_1$$

$$(\mathbf{a}_0 \cdot \mathbf{a}_2) = \cos \Theta_2 = \cos \Theta_1 \cos \gamma + \sin \Theta_1 \sin \gamma \cos \phi \quad (\text{B } 7)$$

$$d\Omega_1 = \sin \Theta_1 d\Theta_1 d\phi_1$$

$$d\Omega_2 = \sin \gamma d\gamma d\phi.$$

Since the integrand does not depend on ϕ_1 , we integrate at once over this variable and get

$$J = 2\pi \int \cosh (\alpha_1 \cosh \Theta_1 + \alpha_2 \cos \Theta_2) F(\sin \gamma) \sin \Theta_1 d\Theta_1 \sin \gamma d\gamma d\phi. \quad (\text{B } 8)$$

The limits of the variables are

$$0 < \Theta_1 < \pi; \quad 0 < \gamma < \pi; \quad 0 < \phi < 2\pi.$$

Now we replace the two variables Θ_1 and ϕ by the following substitution:

$$\cos \Theta_1 = \sin \chi \cos (\psi + \eta(\gamma)) \quad (\text{B } 9)$$

$$\sin \Theta_1 \cos \phi = \sin \chi \sin (\psi + \eta(\gamma))$$

$$\tan \eta(\gamma) = \alpha_2 \sin \gamma / (\alpha_1 + \alpha_2 \cos \gamma)$$

$$\partial(\Theta_1, \phi) / \partial(\chi, \psi) = \sin \chi / \sin \Theta_1$$

whereby the integral takes the form

$$J = 2\pi \int \cosh ([\alpha_1^2 + \alpha_2^2 + 2\alpha_1 \alpha_2 \cos \gamma]^{1/2} \sin \chi \cos \psi) F(\sin \gamma) \sin \gamma d\gamma \sin \chi d\chi d\psi, \quad (\text{B } 10)$$

the integral to be taken between the limits

$$0 < \gamma < \pi; \quad 0 < \chi < \pi; \quad 0 < \psi < 2\pi. \quad (\text{B } 10a)$$

After a final substitution

$$\begin{aligned} \sin \chi \cos \psi &= \cos \mu \\ \cos \chi &= \sin \mu \cos \xi \\ \partial(\chi, \psi)/\partial(\mu, \xi) &= \sin \mu / \sin \chi, \end{aligned} \quad (\text{B } 11)$$

we can integrate over μ and ξ under the integral sign as follows

$$\begin{aligned} & \int_{\mu=0}^{\pi} \sin \mu d\mu \int_{\xi=0}^{2\pi} \cosh ([\alpha_1^2 + \alpha_2^2 + 2\alpha_1 \alpha_2 \cos \gamma]^{1/2} \cos \mu) d\xi \\ &= 4\pi [\alpha_1^2 + \alpha_2^2 + 2\alpha_1 \alpha_2 \cosh \gamma]^{-1/2} \sinh ([\alpha_1^2 + \alpha_2^2 + 2\alpha_1 \alpha_2 \cos \gamma]^{1/2}) \\ &= (-4\pi/\alpha_1 \alpha_2 \sin \gamma) \frac{\partial}{\partial \gamma} \cosh ([\alpha_1^2 + \alpha_2^2 + 2\alpha_1 \alpha_2 \cos \gamma]^{1/2}). \end{aligned}$$

The result

$$J = (-8\pi^2/\alpha_1 \alpha_2) \int_{\gamma=0}^{\pi} \frac{\partial}{\sin \gamma \partial \gamma} \{ \cosh ([\alpha_1^2 + \alpha_2^2 + 2\alpha_1 \alpha_2 \cos \gamma]^{1/2}) \} F(\sin \gamma) \sin \gamma d\gamma \quad (\text{B } 12)$$

may be integrated by parts and we finally obtain

$$\begin{aligned} J &= (8\pi^2/\alpha_1 \alpha_2) \{ 2 \sinh \alpha_1 \sinh \alpha_2 F(0) \\ &\quad + \int_{\gamma=0}^{\pi} \cosh ([\alpha_1^2 + \alpha_2^2 + 2\alpha_1 \alpha_2 \cos \gamma]^{1/2}) dF(\sin \gamma) \} \quad (\text{B } 13) \end{aligned}$$

or in view of (B 6)

$$\begin{aligned} & 2 \sinh \alpha_1 \sinh \alpha_2 \{ \beta_1 \rho(f_1, f_2) - \beta_1(0) \} \\ &= \int_{\gamma=0}^{\pi} \cosh ([\alpha_1^2 + \alpha_2^2 + 2\alpha_1 \alpha_2 \cos \gamma]^{1/2}) d\beta_1(\gamma). \quad (\text{B } 14) \end{aligned}$$

For identical particles, $\alpha_1 = \alpha_2 = \alpha$, we have

$$(\alpha^2 + \alpha^2 + 2\alpha^2 \cos \gamma)^{1/2} = 2\alpha \cos \frac{1}{2}\gamma. \quad (\text{B } 15)$$

Moreover, if we adopt the approximation

$$F(\sin \gamma) = \beta_1(\gamma) = \beta_1(\pi/2) \sin \gamma = -(8/\pi)b \sin \gamma \quad (\text{B } 16)$$

(see text), the integral (B 14) can be expressed in terms of a Bessel function as follows

$$\begin{aligned} 2(\sinh \alpha)^2 \beta_1 \rho(f) &= \int_{\gamma=0}^{\pi} \cosh (2\alpha \cos \frac{1}{2}\gamma) \beta_1(\pi/2) \cos \gamma d\gamma \\ &= \pi \beta_1(\pi/2) I_2(2\alpha) = -8b I_2(2\alpha) \quad (\text{B } 17) \end{aligned}$$

with the standard notation

$$I_2(2\alpha) = -J_2(2i\alpha) = \sum_{n=0}^{\infty} (\alpha^{n+2}/n!(n+2)!) \quad (\text{B } 18)$$

for the Bessel function of order 2. The integral (B 17) is but a variant of the standard integral definition of the Bessel function

$$\pi I_2(2\alpha) = \int_{x=0}^{x=\pi} \cosh(2\alpha \cos x) \cos(2x) dx,$$

the last step being justified by the observation that the integrand remains unchanged when the argument x is replaced by $\pi - x$.

In all cases encountered in the present work, the argument of the Bessel function will be either zero (isotropic solution) or else so large that a few terms of the asymptotic expansion

$$2 I_2(2\alpha) \sim (\pi\alpha)^{-1/2} e^{2\alpha} \left\{ 1 - \frac{3 \cdot 5}{11 \cdot 16\alpha} + \frac{1 \cdot 3 \cdot 5 \cdot 7}{21(16\alpha)^2} - \frac{(-1) \cdot 1 \cdot 3 \cdot 5 \cdot 7 \cdot 9}{31(16\alpha)^3} + \dots \right\} \quad (\text{B } 19)$$

will suffice for computation. The corresponding formula for the mean effective excluded volume is

$$-\beta_1 \rho(f) \sim 8b(\pi\alpha)^{-1/2} \left\{ 1 - \frac{30}{32\alpha} + \frac{210}{(32\alpha)^2} + \frac{1260}{(32\alpha)^3} + \dots \right\}. \quad (\text{B } 20)$$

In the general case $\alpha_1 \neq \alpha_2$, even though we adopt the approximation (B 16), the integral of (B 14) can no longer be expressed in terms of simple known functions. An asymptotic expansion analogous to (B 20) has been obtained by the usual procedure: A new variable t is introduced by the substitution

$$\alpha_1^2 + \alpha_2^2 + 2\alpha_1\alpha_2 \cos \gamma = (\alpha_1 + \alpha_2 - t)^2; \quad (\text{B } 21)$$

the hyperbolic function is approximated by an exponential and the factor

$$dF(\sin \gamma)/dt = \cos \gamma (d\gamma/dt)$$

by an abbreviated power series in t ; finally the range of integration over t is extended to the interval $(0, \infty)$. The following generalization of (B 20) results

$$-\beta_1 \rho(f_1, f_2) \sim 8b_{12}(\alpha_1 + \alpha_2)^{1/2}(2\pi\alpha_1\alpha_2)^{-1/2} \left\{ 1 - \frac{3}{8} \left(\frac{1}{\alpha_1} + \frac{1}{\alpha_2} + \frac{1}{\alpha_1 + \alpha_2} \right) + \frac{15}{128} \left[\frac{8}{\alpha_1\alpha_2} - \left(\frac{1}{\alpha_1} + \frac{1}{\alpha_2} + \frac{1}{\alpha_1 + \alpha_2} \right)^2 \right] + \dots \right\}. \quad (\text{B } 22)$$

By this general technique, it is also possible to deal with the more accurate description of the covolume function given by EQUATION 45 in the text. Some terms involving the factor $\log t$ then occur after the substitution (B 21); but the integrals which correspond to these terms are easily evaluated:

$$\int_0^\infty e^{-t} t^n (\log t) dt = \frac{\partial}{\partial n} \int_0^\infty e^{-t} t^n dt = \Gamma'(n+1).$$

For the present work, however, it did not seem worth while to complete this computation.

Bibliography

- DERJAGUIN, B. 1940. On the repulsive forces between charged colloid particles and on the theory of slow coagulation and stability of lyophobic sols. *Trans. Faraday Soc.* **36**: 203.
- LANGMUIR, I. 1938. The role of attractive and repulsive forces in the formation of tactoids, thixotropic gels, protein crystals and coacervates. *J. Chem. Phys.* **6**: 873.
- ONSAGER, L. 1942. Anisotropic solutions of colloids. *Phys. Rev.* **62**: 558.
- VERWEY, E. J. W., & J. TH. G. OVERBEEK. 1943. *Theory of the Stability of Lyophobic Colloids*. Elsevier Publishing Co. Amsterdam & New York.

THE ATTRACTIONS OF PROTEINS FOR SMALL MOLECULES AND IONS

BY GEORGE SCATCHARD

Department of Chemistry, Massachusetts Institute of Technology, Cambridge, Massachusetts

The number and variety of known compounds between proteins and small molecules are increasing rapidly and make a fascinating story. For instance, there are the compounds of iron, which is carried in our blood plasma by a globulin, two atoms of iron to each molecule of globulin held in a rather tight salt-like binding,¹ which is stored as ferric hydroxide by ferritin much as water is held by a sponge,² and which functions in hemoglobin, four iron atoms in tight porphyrin complexes in each protein molecule. Or, there are many compounds of serum albumin, which was used during the war by many chemists, most of whom found at least one new compound. This molecule, which has about a hundred carboxyl radicals, each of which can take on a proton, and about the same number of ammonium radicals, each of which can dissociate a proton, has one single radical which combines with mercuric ion so firmly that two albumin molecules will share one mercury atom if there are not enough to go around.³

At the present stage of rapid growth of known compounds, it seems more profitable for me to make no attempt to catalogue the various classes of compounds, but to discuss the general principles involved, in the hope that this will make more useful the information which is accumulating so rapidly from so many laboratories.

We want to know of each molecule or ion which can combine with a protein molecule, "How many? How tightly? Where? Why?" The answer to the first two questions, and sometimes to the third, can be furnished by the physical chemist, but he will often need to team with an organic chemist to determine the effect of altering specified groups to find if they are reactive. The determination of function is a complicated problem which may be the business of the physiologist or physiological chemist. But the answers to both of the more complicated problems will depend on the answers to the simpler questions, "How many?" and "How tightly bound?"

If the various groups on a protein molecule act independently, we can apply the law of mass action as though each group were on a separate molecule,⁴ and the strength of binding can be expressed as the constant for each group. Often, a single constant will express the behavior of several groups. If the constants are widely spread, as those for the reaction of hydrogen ion with carboxylate ions, with imidazoles and with amines, the interpretation is simple. If the separation is less, it is very difficult to distinguish the case of different intrinsic affinities from the case of interaction among the groups.

We know that such interaction occurs in simple molecules in which a reac-

tion has equal probability of happening at various points on a molecule. Reaction at one of these points may make it much more difficult for the reaction to occur at another point, as in the dibasic carboxylic acids,⁵ or it may make a second reaction much easier, as in the reaction of ammonia with silver ion.⁶ There may be an effect of the medium which can be interpreted by an activity coefficient, but there may also be a residue which is independent of the medium. There may be an electrostatic effect, but there may also be additional effects which cannot be explained by any simple electrostatic theory.

Independent action of the groups means that the change in free energy for the reaction of the protein with ν small molecules is made up of the statistical entropy terms plus a term proportional to ν . The simplest extension is to add another term proportional to ν^2 . This extension is particularly important since it is sufficient to account for the Debye-Hückel approximation of electrostatic interaction in a medium of unchanging dielectric constant and ionic strength, or to account for non-electrostatic interaction with random distribution.

If the initial probability of reaction is the same at each of n points, the change in free energy (ΔF), for the reaction $P_0 + \nu A = PA_\nu$, is given by

$$(\Delta F)_\nu/RT = \ln c_\nu/c_0c_A^\nu + \ln \nu!(n-\nu)!/\nu! - \nu \ln k + w\nu^2 \quad (1)$$

in which RT has its usual significance, c_0 , c_ν , and c_A are the concentrations of P_0 , PA_ν , and A , k is the intrinsic constant for the reaction at a single group, $\nu!$ is ν factorial and w is the coefficient of ν^2 . The average association

$$\bar{\nu} = \sum_{\nu=0}^n c_\nu \nu / \sum_{\nu=0}^n c_\nu$$

$$\bar{\nu} = \frac{\sum_{\nu=0}^n \frac{n!}{\nu!(n-\nu)!} (kc_A)^\nu e^{-w\nu^2} \nu}{\sum_{\nu=0}^n \frac{n!}{\nu!(n-\nu)!} (kc_A)^\nu e^{-w\nu^2}} \quad (2)$$

The calculation of $\bar{\nu}$ by this equation is straightforward, and may be extended to the case of more than one constant by addition of the respective $\bar{\nu}$'s taking into consideration the possibility that the $w\nu^2$ terms may become more complicated. This method has been used by Cannan, Kibrick, and Palmer⁷ for the titration of fifty-one carboxyls, five imidazoles, and twenty-three amines in ovalbumin, and by Klotz, Walker, and Pivan⁸ for twenty-two sulfathiazole groups reacting with serum albumin. If the total number of groups is large, however, the method is very tedious, and if the total number is unknown it is practically unusable.

Linderstrom-Lang⁹ attempted to sum the series in the paper in which he made the first application of the Debye theory to titrations of proteins. He obtained the effect on the straight middle portion of the curve of $\bar{\nu}$ versus pH, but did not extend it further. Cannan, Kibrick, and Palmer⁷ used the com-

plete expression. Putzeys and Bouckaert¹⁰ derived the complete expression with very complicated mathematics. The solution for a very large number of groups is so simple and holds so well for a moderately large number of groups that it is worth while to present it free from any non-essentials.

From EQUATION 1, we find that the ratio of the concentration of all species with ν molecules of A combined with one of protein to the concentration of those with $\nu - 1$ molecules is

$$\frac{c_\nu}{c_{\nu-1}} = \frac{n+1-\nu}{\nu} k c_A e^{-w(2\nu-1)}. \quad (3)$$

If the titration is to spread over only a few powers of ten, $n^2 w$ must be finite and only moderately large. Thus, $w(2\nu-1)$ must be very small when n is large. Therefore, $c_\nu/c_{\nu-1}$ will be unity at the same value of $\bar{\nu}$ as for an ideal solution,* for which $w = 0$ and $k c = \bar{\nu}/(n - \bar{\nu})$, and $\nu = \bar{\nu} + \bar{\nu}/n$: substituting this value in EQUATION 3 and transposing yields

$$k e^w c_A = \frac{\bar{\nu}}{n - \bar{\nu}} e^{2(1+1/n)w\bar{\nu}}$$

$$k' c_A = \frac{\bar{\nu}}{n - \bar{\nu}} e^{2w'\bar{\nu}} \quad \ln k' c = \ln \frac{\bar{\nu}}{n - \bar{\nu}} + 2w'\bar{\nu} \quad (4)$$

in which $k' = k e^w$ and $w' = \left(1 + \frac{1}{n}\right)w$. Rather than trying to prove that

this is the limiting expression for very large values of n , let us see how bad it is for very small values.

We find that titration curves often have shapes not unlike that for $w = 0$, which is shown as curve 4 in FIGURE 1, except that they are spread about

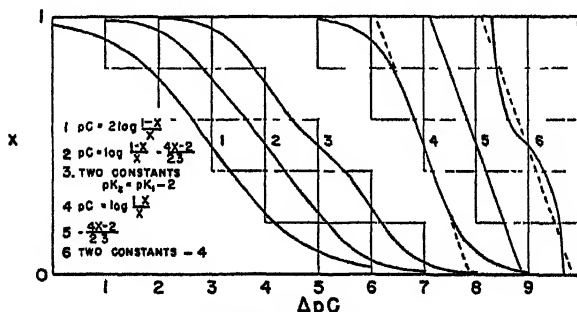


FIGURE 1. Titration Curves.

twice as far on the pH axis. How can the ideal curve be warped to give this spread? Although the central portion of the curve is nearly linear, it is not

* We note that, if w is so large that reactions at different points are completely separated, so that PA_1 and PA_2 are the only forms of the protein present, $\nu = \bar{\nu} + \frac{1}{2}$. The difference, $\bar{\nu}/n - \frac{1}{2}$, is to be compared to n . Moreover, the correct value is much nearer $\bar{\nu}/n$ than to $\frac{1}{2}$ for any probable values of w .

permissible to rotate the curve about its center because the asymptotes must remain horizontal with $x = 0$ and $x = 1$. If the reactions at half the points have an intrinsic dissociation constant different from those at the other half, we divide curve 4 into two equal parts, pull them apart horizontally, and then add. Curve 3 shows the result if k'' is one-hundredth of k' . If EQUATION 4 is valid, each point should be displaced horizontally by an amount proportional to its perpendicular distance from the midpoint. Curve 2 corresponds to EQUATION 4 with $w'n = 2$. It is also possible to displace each point by an amount proportional to its horizontal distance from the midpoint, which is equivalent to changing the scale of abscissae. In curve 1, each point has double the horizontal displacement of curve 4. When the displacement of curve 4 is divided by an integer, the resulting curve corresponds to reaction of that integral number of molecules with one protein without any intermediate forms. I can find no physical explanation for this type of curve for displacement greater than that of curve 4, but I have included it because this expression is often used. Curves 4, 5, and 6 represent the difference in horizontal displacement of curves 1, 2, and 3 from 4, and the broken lines with 4 and 6 are repetitions of 5.

FIGURE 2 shows the differences in \bar{v}/n of curve 3 of FIGURE 1 from curve 2

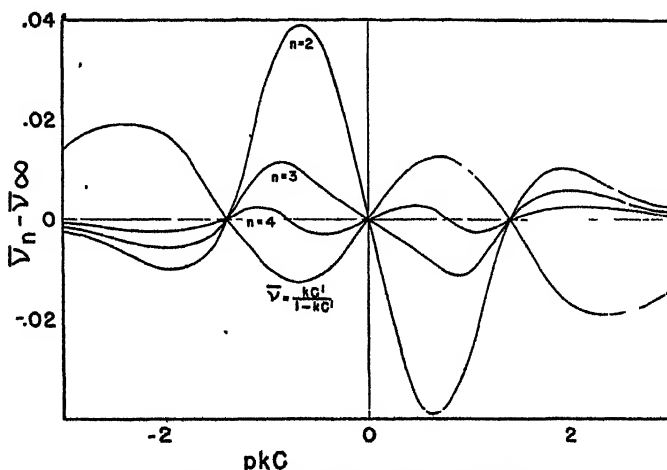


FIGURE 2. Effect of Finite Number of Groups.
Deviations from $pkC = \log \frac{1-x}{x} - \frac{4x-2}{2.3}$.

at the same value of pkC . The maximum difference is 3.9 percent, and the difference is zero when kC is 0.04, 1, or 25. FIGURE 2 also shows the difference from curve 2 of a curve of the type of curve 1, and of the curves for $n = 3$ and $n = 4$ which intersect curve 2 at these same points as curve 3. For three groups, the maximum deviation is 1.2 percent; for four groups it is 0.3 percent, and the deviation is also zero for $kC = 0.1$ and 10. Thus, four is

practically infinity within the accuracy of most measurements if the curve is spread twice the width of curve 4, FIGURE 1.

For the Debye-Hückel approximation for a charge spread uniformly over the surface of a sphere of radius b which excludes small ions to a radius a ,

$$w = \frac{\epsilon^2 z^2}{2DkT} \left(\frac{1}{b} - \frac{\kappa}{1 + \kappa a} \right) \quad (5)^{*},^{11}$$

in which D is the dielectric constant of the medium, k is Boltzmann's constant, T the absolute temperature, ϵ the electronic charge, z the valence of the small molecule, and κ has its usual significance in the Debye theory, and the net valence of the protein replaces ν in w^2 .

The expression for a discrete distribution of charges within a spherical molecule is given by Kirkwood,¹² as modified by Kirkwood and Westheimer.¹³ The greatest difference from the Debye-Hückel expression is that b is a function of the charge distribution and dielectric constant of the protein molecule, and is such a function of the dielectric constant of the medium that $1/Db$ does not vanish for an infinite dielectric constant. Kirkwood's expression also gives a more complicated variation with ionic strength, but this will be so smeared by non-electrostatic effects that, in the present state of our knowledge, it will be sufficient to use this equation with the conditions that b need not be too closely related to the size of the molecule, and a is always greater than b . These expressions are all limited to very small protein concentrations. The effect of increasing concentration is discussed by Scatchard, Batchelder, and Brown.¹⁴ Usually we do not know enough about the protein to justify considering second approximations.

At times, I have been troubled by the fact that the probability of reaction is not the same for all points once reaction has occurred at one of them. I know that I have not been alone in this worry, so it has seemed worth while to consider the electrostatic effects for two simple distributions: a regular tetrahedron and a cube at zero ionic strength in a medium of the same dielectric constant as the large molecule. The results are shown in FIGURE 3. Putting on the first small molecule will require no electrostatic work, but for each additional one there will be work proportional to the sum of the reciprocal distances to each charge already there.

For the regular tetrahedron, the distances are all the same and there is one form with no charge, four with one charge, six with two, four with three, and one with four charges. The works are proportional to 0, 1.7, 5.2, 10.3, or to $0.86 \nu(\nu - 1)$. The works are normalized so that the work for the fully charged cube is $\nu(\nu - 1)$. For the smaller charges, the number of each form and the work is listed below the model of the form. The differences from $\nu(\nu - 1)$ are so small for most of the molecules that my worries have been dissipated.

* The limit at very small values of κ is $(\epsilon^2 z^2 / 2DkT) / b$, and at very large values of κ it is $(\epsilon^2 z^2 / 2DkT) (1/b - 1/a)$. Two parameters, a and b , are necessary to keep the latter limit different from zero.

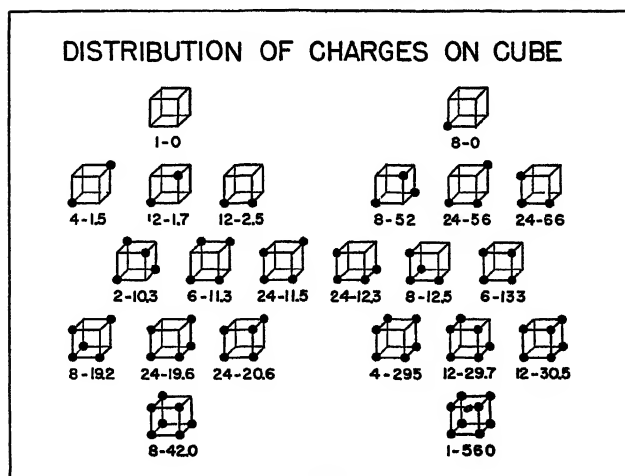
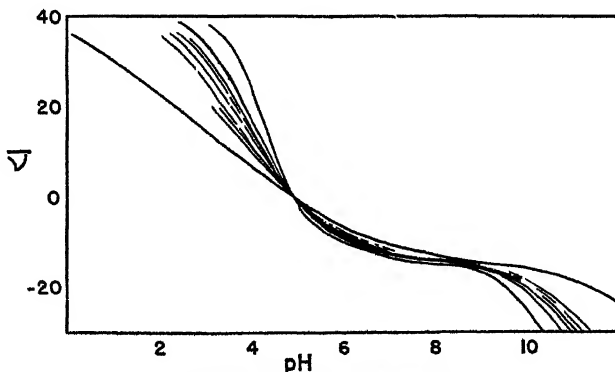


FIGURE 3.

There seems to be but little known of negative values of w for protein compounds. However, we might expect cases when the van der Waal's attraction of long chains should overcompensate electrostatic repulsion just as it does in soap micelles. When mw equals -2 , the slope of \bar{v} vs. $\log c_A$ becomes infinite, and for larger values there is a sudden jump from small values to large. The electrophoresis of albumin and decanyl sulfate by Putnam and Neurath¹⁵ probably indicates an effect of this kind. The first compound may represent saturation with charged ends toward the protein, while the second may have a reversed layer, giving a soap micelle wrapped around the protein molecule.¹⁶

FIGURE 4 shows the titration of ovalbumin with HCl or KOH in the presence of KCl in various amounts by Cannan, Kibrick, and Palmer.⁷ The

FIGURE 4. Titration of Ovalbumin.⁷

extent of combination necessary to explain these deviations as pure Donnan effects. The value of \bar{v} falls regularly as the valence becomes less positive, jumps suddenly near the isoionic point, and then falls again steadily. A glimpse at FIGURE 8 will explain the behavior. The points are B from the equation

$$PV^{\circ} = RTm_2(1 + B\bar{W}_2m_2)$$

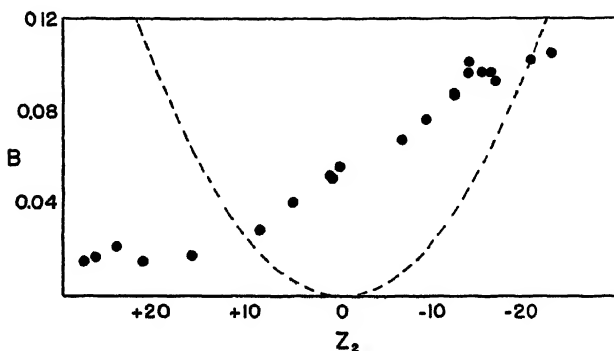


FIGURE 8. Serum Albumin and Chloride Ion ¹⁴

in which P is the osmotic pressure, V° the volume containing a kilogram of water, m_2 the molal concentration of protein, and \bar{W}_2 its molecular weight. The broken line is the Donnan value. The extent of combination calculated in FIGURE 7 is the horizontal displacement necessary to bring the point to the Donnan line, and the break near the isoionic point separates those points which are moved to the right branch from those which are moved to the left. Taking into account the effect of the combination indicated by the salt distribution makes the effect of albumin on its own activity coefficient more symmetrical, but it does not reduce much the maximum effect, which must be taken into account.

Recently, we have been studying the combination of chloride ion and of thiocyanate ion with human serum albumin by a procedure like that of Klotz,⁸ except that the concentration is determined by conductance. This necessitates measurements without added salt and therefore at varying ionic strength. The experimental results will be reported in a later paper, but I want to discuss here the method of treatment which we developed for them.

For the acid titrations, the maximum binding capacity can be approached closely with relatively low concentrations of acid or base. For the weaker associations, which we are considering now, the average amount bound is still increasing rapidly when only a small fraction of the small molecules are combined. This makes the determination of the maximum binding less certain, and our task is to reduce that uncertainty as far as possible.

Recent usage has been to invert the law of mass action solved for \bar{v} to give

$$\frac{1}{\bar{v}} = \frac{1 + kc}{knc} = \frac{1}{n} + \frac{1}{knc} \quad (6)$$

to plot $1/\bar{v}$ against $1/c$, to draw the best straight line and call its intercept $1/n$ and its slope $1/kn$. This has the disadvantage of concealing deviations from the ideal laws, and of tempting straight lines where there should be curvature.

I have preferred to start with the mass action law solved for c :

$$\bar{v}/(n - \bar{v}) = kc$$

and multiply by $(n - \bar{v})/c$ to give

$$\bar{v}/c = k(n - \bar{v}). \quad (7)$$

Plotting \bar{v}/c against \bar{v} again gives a straight line if k is constant. The intercept on the \bar{v}/c axis is kn , the classical first association constant, and the intercept on the \bar{v} axis is n . This plot shows immediately how great is the extrapolation necessary to determine these quantities.

Curvature may indicate different intrinsic constants or deviations from independent probabilities. In the latter case, we may alter EQUATION 4 to give

$$\bar{v}e^{2w'\bar{v}}/c = k'(n - \bar{v}) \quad (8)$$

and plot $\bar{v}e^{2w'\bar{v}}/c$ against \bar{v} . Sometimes w' may be calculated theoretically, or an approximate value may be determined empirically. It is not necessary to straighten the line if the correction is good enough to determine the intercepts. Even if there are different intrinsic constants, the two intercepts are still the classical first association constant and the total number of groups.

As an example, we show the titration of ovalbumin of Cannan, Kibrick, and Palmer⁷ with 32 g. protein per liter and no added salt. In FIGURE 9, the

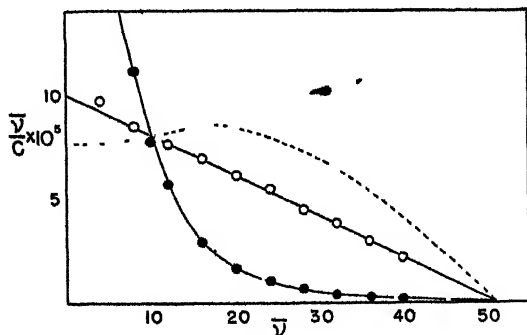


FIGURE 9 Titration of Ovalbumin⁷

full circles are the measured values of \bar{v} assuming that \bar{v} is ten for the isoionic protein. The open circles are corrected by the equation corresponding to

the authors' assumption that the correction is 0.8, the Debye value for $b = 27.5 \text{ \AA}$ and $a = 29.5 \text{ \AA}$, which gives

$$2w/2.3 = 0.084[1.073 - 9.68\sqrt{\mu}/(1 + 9.68\sqrt{\mu})]$$

in which μ is the ionic strength. The straight line corresponds to their values of $n = 51$ (obtained from special experiments) and $\log k = 4.29$. The curve is obtained from this line by making the correction in reverse. The measured value for the most dilute point falls off the scale. The corrected value begins to show the effect of the imidazole groups, which are not counted in \bar{p} .^{*} The broken line shows the effect of correcting the lower curve by the full Debye value for the dimensions chosen by the authors, that is, with w 1.25 times that for the straight line. Although it is obviously overcorrected, the values of $(n - \bar{p})$ and of $k\bar{p}$ at the isoionic point could be obtained from it with very fair accuracy.

FIGURE 10 shows the results of Klotz, Walker, and Pivan⁸ on bovine serum

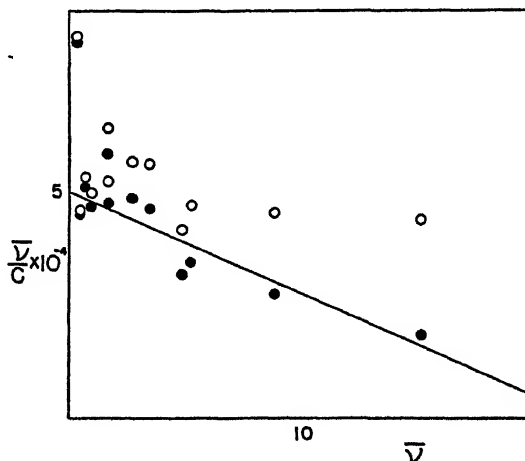
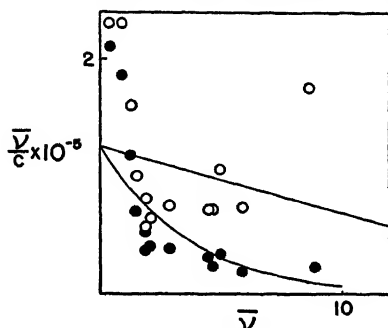
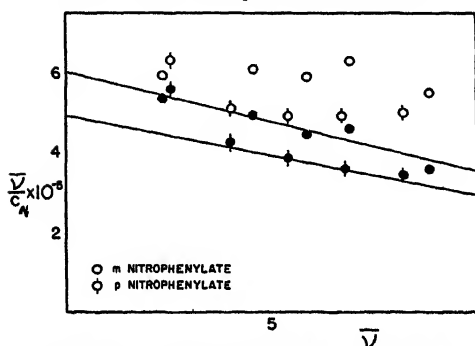
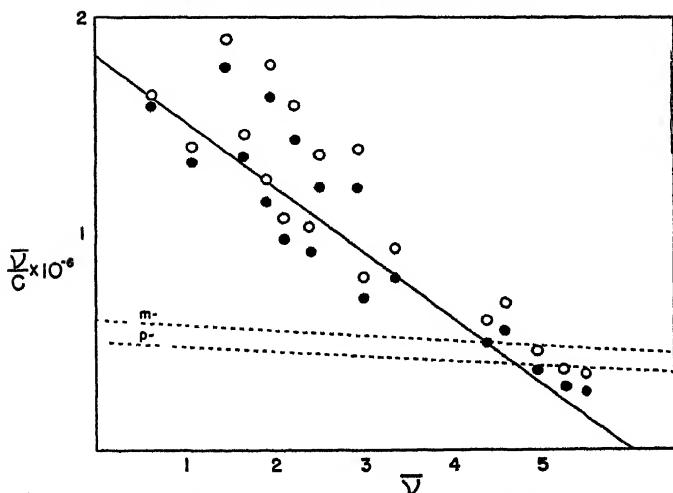


FIGURE 10. Serum Albumin and Methyl Orange.⁸

albumin and methyl orange. Again, the filled circles are the measured values and the open circles are corrected for electrostatic effects using the dimensions which the authors used for sulfathiazole. The straight line is that determined by the values of n and k given by the authors. It is not certain that these measurements should be corrected for electrostatic effects. They are made in 0.1 M phosphate buffer, and it is possible that the methyl orange replaces phosphate ion instead of reacting with uncombined albumin. The extrapolation of these results is much less certain than that in FIGURE 9.

FIGURE 11 shows their results for serum albumin and sulfathiazole. The correction is the same as that made by the authors and the straight line is

^{*} The correction is made for $\bar{p} = 10$, so that there is none at the isoionic point, and the value of k should be determined as $\bar{p}/c(n - \bar{p})$ for the isoionic point.

FIGURE 11. Serum Albumin and Azosulfathiazol.⁸FIGURE 12. Serum Albumin and m- and p-Phenylate.²⁰FIGURE 13. Serum Albumin and o-Nitrophenylate.²⁰

determined from their values of n and k . The curve through the experimental points is made by applying the correction in reverse to the straight

line. If the reaction is the displacement of a univalent ion by a bivalent one, the correction should be only the fourth root of that used.

FIGURES 12 and 13 show the results of Teresi and Luck²⁰ for bovine serum albumin and some nitrophenolates. Again, the filled circles represent the measured values, the open circles are corrected for electrostatic effects as in FIGURE 10, and the lines are determined from the values of k and n given by the authors. These reactions may also be displacements of buffer anions so the electrostatic correction may be improper.

Although the precision of extrapolation is not very great, the results are quite sufficient to show the difference between the orthophenolate on one hand and the meta- and paraphenolates on the other. For the orthophenolate, k is much larger and n much smaller. The authors find smaller values of n for many orthonitrophenols, and they attribute the difference to steric hindrance. It is not surprising that an ortho-nitro group favors association at points where the steric hindrance does not interfere.

Much of the difference between the precision of extrapolation in FIGURE 9 for hydrogen ion and in the subsequent figures for anions depends upon the much greater magnitude of the association constants for the acid titration. However, these figures should show the great importance of the greatest possible precision over the widest possible range in order that these curves may be extrapolated accurately to the intercepts which tell us "how many" and "how tightly bound."

Bibliography

1. SORGENOR, D. M., B. A. KOECHLIN, & L. E. STRONG. J. Clin. Inv. In press.
2. MICHAELIS, L. 1947. Advances in Protein Chemistry 3: 53.
3. HUGHES, W. L. 1947. J. Am. Chem. Soc. 69: 1836.
4. VON MURALT, A. L. 1930. J. Am. Chem. Soc. 52: 3518.
5. HAMMETT, L. P. 1940. *Physical Organic Chemistry*: 201. McGraw Hill. New York.
6. BJERRUM, N. 1941. *Metal Amine Formation in Aqueous Solution*. P. Haase and Son. Copenhagen.
7. CANNAN, R. K., A. KIBRICK, & A. H. PALMER. 1941. Ann. N. Y. Acad. Sci. 41: 243.
1942. Chem. Rev. 30: 395.
8. KLOTZ, I. M., F. M. WALKER, & R. B. PIVAN. 1946. J. Am. Chem. Soc. 68: 1486.
9. LINDERSTROM-LANG, K. 1924. C. R. Lab. Carlsberg 15 (7).
10. PUTZEYS, P., & L. BOUCKAERT. 1942. Medd. Kon. Vlaamsche A. Wet., L. & S. K., van Belgie 4 (3).
11. SCATCHARD, G. 1927. Trans. Faraday Soc. 23: 454.
12. KIRKWOOD, J. G. 1934. J. Chem. Phys. 2: 351.
13. KIRKWOOD, J. G., & F. H. WESTHEIMER. 1938. J. Chem. Phys. 6: 506.
14. SCATCHARD, G., A. C. BATCHELDER, & A. BROWN. 1946. J. Am. Chem. Soc. 68: 2320.
15. PUTNAM, F. W., & H. NEURATH. 1944. J. Am. Chem. Soc. 66: 1992.
16. PARKHURST, K. G. A., & R. C. N. SMITH. 1944. Trans. Faraday Soc. 40: 565.
17. STEINHARDT, J., & M. HARRIS. 1940. J. Res. N. B. S. 24: 335.
18. STEINHARDT, J. 1941. Ann. N. Y. Acad. Sci. 41: 287.
19. STEINHARDT, J., C. H. FUGITT, & M. HARRIS. 1942. J. Res. N. B. S. 28: 201.
20. TERESI, J. D., & J. M. LUCE. 1948. J. Biol. Chem. 173: 653.

DETERMINATION OF THE DISSOCIATION CONSTANTS OF WEAK ELECTROLYTES IN SALT SOLUTIONS

BY MARTIN KILPATRICK

Department of Chemistry, Illinois Institute of Technology, Chicago, Illinois

In connection with kinetic studies, it is often essential to have a fairly accurate knowledge of the hydrogen ion concentration of aqueous and non-aqueous solutions in the presence of appreciable concentrations of electrolytes. There now exist precise values of the thermodynamic dissociation constants of a number of carboxylic acids in aqueous solution, and in some cases the dissociation constants are known over a considerable range of concentration for a few electrolytes. These dissociation constants have been determined by five general methods: (1) conductance; (2) electromotive force methods; (3) colorimetry; (4) kinetic measurements; and (5) solubility.

It is the purpose of this paper to review the use of the kinetic method for determining the dissociation constants of monobasic acids and to point out the difficulties and assumptions involved when the concentration of electrolytes is high.

In the early days of the Arrhenius theory, Ostwald and others used the kinetic method to obtain the ratio of the dissociation constants of acids. Essentially, the method consisted in obtaining the velocity constant for the inversion of sucrose or the hydrolysis of methyl acetate in solutions containing equivalent concentrations of the acids in question. The ratio of the velocity constants was taken as the ratio of the dissociation constants. These results were often combined with dissociation constants obtained from conductance calculations based on the assumption that $\Lambda/\Lambda_{\infty} = \alpha$ and the Ostwald Dilution Law. For example, Ostwald¹ determined the dissociation constant of dichloroacetic acid from conductance measurements at 25° C. as 0.0514 and then determined the ratio of this dissociation constant to that of trichloroacetic acid by a comparison of the rates of inversion of sucrose for the two acids and reported a value of 1.2 for the dissociation constant of trichloroacetic acid. We shall see later that there is no reliable value of the dissociation constant of trichloroacetic acid in the literature at the present time. In using the kinetic method to determine hydrogen ion concentration, Ostwald assumed that the rate of reaction was directly proportional to the hydrogen ion concentration and neglected any electrolyte effects on the rate of reaction.

With the introduction of empirical rules for activity coefficients^{2,3} and of the Debye-Hückel equation,⁴ together with the solution of the conductance problem,⁵ many of the difficulties of methods 1, 2, 3 and 5 for determining

dissociation constants disappeared. A step forward was taken in the use of the kinetic method by introduction of the equation for reaction velocity.⁶

For a reaction between *A* and *B*, this equation is

$$v = k^{\circ} C_A C_B \frac{f_A f_B}{f_{(A, B)}} \quad (1)$$

and the application of the Debye-Hückel theory to this equation for reaction between ions yields, for water at 25° C.,

$$\log k = \log k^{\circ} + Z_A Z_B \sqrt{\mu} \quad (2)$$

an equation which has proved most successful in explaining electrolyte effects in reactions between ions in the range of concentration where the Debye-Hückel limiting law holds. For aqueous solutions at ionic strengths above 0.01, where the limiting law of Debye and Hückel is no longer applicable, the introduction of a salting-out term has been fairly successful. This equation may be written as follows:

$$\log k = \log k^{\circ} + Z_A Z_B \sqrt{\mu} + B\mu \quad (3)$$

where $B = [-\beta_A - \beta_B + \beta_{(A, B)}]$. When EQUATION 3 is applied to a reaction between an ion and a non-electrolyte, the second term drops out, and EQUATION 3 for hydrogen ion catalysis takes the form

$$\log k_{H_3O^+} = \log k^{\circ}_{H_3O^+} + BC \quad (4)$$

where $k_{H_3O^+}$ represents the velocity constant for molar hydrogen ion at the concentration *C* of uni-univalent salt, and $k^{\circ}_{H_3O^+}$ is the velocity constant for molar hydrogen ion at zero electrolyte concentration. It is the purpose of the present paper to examine the application of EQUATION 4 to reactions between an ion and a non-electrolyte and, in particular, to reactions where the ion is the hydrogen ion in aqueous solution.

An examination of the results in the literature indicates that the magnitude of the electrolyte effect varies greatly. TABLE 1 gives the *B* values for a

TABLE 1
ELECTROLYTE EFFECT $B = \log \frac{k_{H_3O^+} - \log k^{\circ}_{H_3O^+}}{C}$

Reaction	HCl	HClO ₄	HNO ₃	pC ₆ H ₄ CHSO ₃ H	C ₆ H ₅ SO ₃ H
Hydration of isobutene	0.44	0.47	0.21	0.30	
Hydrolysis of ethylal	0.35				
Hydrolysis of dimethyl acetal	0.31				
Hydrolysis of sucrose	0.21	0.29	0.17		0.25
Hydrolysis of ethyl acetate	0.078				

number of reactions, and it is to be noted that these values vary from reaction to reaction and from electrolyte to electrolyte for the same reaction.

As an illustration, B varies from 0.44 to 0.078 for the electrolyte hydrochloric acid for five different reactions, and it varies from 0.47 for perchloric acid to 0.21 for nitric acid in the case of the hydration of isobutene.⁷

All of these reactions are accepted as cases of specific hydrogen ion catalysis. The effect of electrolytes at constant hydrogen ion concentration has also been studied in the case of the hydrolysis of acetals,⁸ and the large and varying magnitude is shown by TABLE 2, which gives the percentage salt effect at

TABLE 2
ELECTROLYTE EFFECT OF DIFFERENT SALTS ON THE HYDROLYSIS OF DIETHYL ACETAL

<i>Salt</i>	% Increase in $k_{H_3O^+}$ at 0.1 <i>M</i>
NaClO ₄	14.2
LiCl	12.8
NaCl	13.2
KCl	12.6
LiNO ₃	9.3
NaNO ₃	10.2
KNO ₃	7.5
C ₆ H ₅ SO ₃ Na	5.3
pC ₆ H ₄ CH ₃ SO ₃ Na	5.0

0.1 molar for nine electrolytes. These results are to be compared to 2–4 per cent effects in the case of the hydrolysis of ethyl acetate. Olson and Tong⁹ have considered this problem and, in the cases where the activity coefficient of the substrate is known, have concluded that the variation of the ratio of the activity coefficient of the hydrogen ion to that of the activity coefficient of the critical complex, with electrolyte concentration, would have to be much greater than expected for the ratio of univalent ions in order to account for the magnitude of the electrolyte effect. These authors suggest that the ions of electrolyte polarize the water molecules and this orientation of the water molecules changes the specific rate of the reaction.

Whatever the explanation, the magnitude of the electrolyte effect necessitates certain assumptions in the use of the kinetic method in the determination of hydrogen ion concentration. Riesch and Kilpatrick¹⁰ determined the dissociation constant of benzoic acid in nine solvent salts by the following method. The hydrolysis of diethyl acetal was chosen as a reaction giving a convenient rate at 25° at a hydrogen ion concentration of 1×10^{-4} moles per liter. The reaction was calibrated at 0° in the presence of the salts with 0.01 molar strong acid as the catalyst. In all cases, the electrolyte was in large excess and $k_{H_3O^+}$ was determined at 0° C. to avoid the use of low concentration of strong acids. The calibration curves for the reaction were transferred to 25° on the assumption that the energy of activation for the reaction, determined from the temperature coefficient k_{35}/k_{25} , was independent of temperature and electrolyte concentration. The kinetic measurements with benzoic acid-benzoate buffers in the presence of an excess of salt

permitted the determination of the hydrogen ion concentration and calculation of the dissociation constant. The K_c/K_a ratios, calculated using the thermodynamic dissociation constant obtained from conductance measurements,¹¹ are given in TABLE 3. The K_c/K_a ratios vary from electrolyte to

TABLE 3
RATIOS OF THE CONCENTRATION TO THERMODYNAMIC DISSOCIATION FOR BENZOIC ACID,
 K_c/K_a

μ	KCl	NaCl	LiCl	NaClO ₄	KNO ₃	NaNO ₃	LiNO ₃	C ₆ H ₅ SO ₃ Na	pC ₆ H ₄ CH ₂ SO ₃ Na
0.10	1.61	1.68	1.72	1.65	1.64	1.65	1.62	1.60	1.57
0.20	1.78	1.82	1.86	1.72	1.76	1.79	1.79	1.74	1.62
0.30	1.81								
0.40		1.86	2.04	1.66	1.82	1.86	1.97	1.74	1.52
0.50	1.82							1.70	
0.60		1.89	2.04	1.62	1.82	1.86	1.95		1.30
0.80	1.77	1.88	2.04	1.53	1.77	1.83	1.93	1.49	1.05
1.0	1.72	1.81	2.03	1.41	1.72	1.77	1.90	1.31	0.87
1.5	1.52	1.57	1.93	1.18		1.58	1.75	0.84	
2.0	1.29	1.38	1.71	0.97		1.43	1.56		
2.5	1.11	1.07	1.52			1.27			
3.0	0.80	0.77	1.20	0.59		1.08	1.13		

electrolyte, and where comparisons with other methods have been made¹² the results are reasonably concordant. A further examination of the assumption of the constancy of the measured energy of activation showed a decrease of 33 cal. per degree rise in temperature for diethyl acetal.¹³

In the case of acids of higher dissociation constant (greater than 10^{-3}), the reaction can be calibrated with approximately 0.01 molar hydrochloric acid, and the dissociation constant determined kinetically without involving the assumptions of the constancy of the energy of activation in the Arrhenius equation. However, halogen-substituted acids such as monochloroacetic acid are not stable at low hydrogen ion concentrations, and the increased accuracy resulting from the use of a buffer solution is lost. For example, kinetic measurements on the hydrolysis of ethylene acetal in monochloroacetic acid solution yield values of the dissociation constant which, upon extrapolation, give 1.413×10^{-3} ¹⁴ for the thermodynamic dissociation constant of monochloroacetic acid at 20°. This compares favorably with values of 1.49 and 1.44×10^{-3} from measurements¹⁵ of cells with and without liquid junction at 18° C. At 25°, the value from measurement of cells without liquid junction¹⁵ is 1.379×10^{-3} , 1.400×10^{-3} from conductance measurement,¹⁶ and 1.40×10^{-3} from colorimetric measurements.¹⁷

In the case of trichloroacetic acid, the observed velocity constant for the

hydration of isobutene is equal to that for nitric acid.⁷ On the assumption that the kinetic electrolyte effect of trichloroacetate is equal to that of nitrate ion, trichloroacetic acid would necessarily be considered as strong an acid as nitric acid. The dissociation constant, calculated on the assumption that the electrolyte effect of perchlorate is the same as of trichloroacetate, is 1.0. TABLE 4 summarizes the calculation of the dissociation constant from the

TABLE 4

DISSOCIATION CONSTANT OF TRICHLOROACETIC ACID FROM INVERSION OF SUCROSE

CCl_3COOH <i>m/l</i>	H_2O^+ <i>m/l</i>	K_a	K_a/K_a	K_a
0.5	0.362	0.949	1.86	0.508
2.0	0.932	0.812	1.80	0.451
4.0	1.35	0.688	1.64	0.420
	1.41	0.768	1.62	0.474

Assumptions (a) Reaction catalyzed solely by hydrogen ion.

(b) Electrolyte effect same as perchloric acid

$$\log k_{\text{H}_2\text{O}^+} = 3.560 + 0.290 \text{ C}_{\text{HClO}_4}$$

(c) Activity coefficients vary with ionic concentration as those for acetic acid in sodium chloride solutions.

inversion of sucrose¹⁸ on the same assumptions, which gives approximately 0.5, and recalculation of the kinetic data for the hydrolysis of ethyl acetate¹⁹ gives approximate agreement with the results from the hydration of isobutene. However, a calculation neglecting the effect of ionic atmosphere on the mobility of the ions, using the conductance data of Ostwald, yields a value of the dissociation constant of trichloroacetic acid which is almost a power of 10 lower, after correction for activity coefficients. The conductance data are not accurate enough to attempt a calculation using the procedure of Onsager.⁵ Hall²⁰ reports a value of 0.2. Baughan²¹ rightly claims there is no reliable value of the dissociation constant of trichloroacetic acid in the literature.

In view of these discrepancies, it is of interest to compare the kinetic determination of the dissociation constant of dichloroacetic acid using more than one reaction. TABLE 5 gives the dissociation constant of dichloroacetic acid from kinetic measurements with ethylene acetal which are to be compared with 0.043 from the kinetic determination using the hydration of isobutene and 0.0332 reported by Harned and Hawkins²² from kinetic measurements on the hydrolysis of ethyl acetate. Our own analysis of these data, given in TABLE 6, yields a somewhat higher value.

The kinetic results are to be compared with 0.0514 reported by Ostwald. A recalculation of part of the conductance data, taking into account the change of mobility and activity coefficients with ion concentration, yields a value of 0.046 for the thermodynamic dissociation constant at 25° C.; but too

TABLE 5

DISSOCIATION CONSTANT OF DICHLOROACETIC ACID IN SOLUTIONS OF SODIUM DICHLOROACETATE AND SODIUM CHLORIDE FROM KINETIC MEASUREMENTS WITH ETHYLENE ACETAL AT 25° C. (HILLENBRAND AND KILPATRICK)

H_2O^{+*} <i>m/l</i>	$\text{CHCl}_2\text{COO}^-$ <i>m/l</i>	CHCl_2COOH <i>m/l</i>	K_o $\mu = 0.5$	K_a^{**}
0.02559	0.1460	0.03973	0.0940	0.0497
.02465	.1451	.04067	.0879	.0465
.03062	.2723	.1005	.0829	.0438
.03226	.3917	.1628	.0776	.0410
.03204	.5135	.2290	.0718	.0380

* $[\text{H}_2\text{O}^+] = k_{\text{obs}}/k_{\text{H}_2\text{O}^+}$. $k_{\text{H}_2\text{O}^+}$ was obtained from the HCl-NaCl experiment for which $[\text{H}_2\text{O}^+] = 0.025$ m/l and $k_{\text{obs}} = 0.02952$.

** $K_a = K_o/1.892$, where 1.892 is the activity coefficient factor corresponding to an ionic strength of 0.51 to 0.52.

TABLE 6

DISSOCIATION CONSTANT OF DICHLOROACETIC ACID FROM THE HYDROLYSIS OF ETHYL ACETATE

(All solutions 0.202 molal in CHCl_2COOH .)

NaCl <i>m/l</i>	K_o	K_o/K_a	K_a
0.1987	0.0686	1.82	0.0377
.4942	.0691	1.88	.0368
.9790	.0625	1.75	.0357
1.920	.0486	1.38	.0352

Assumptions: (a) Reaction catalyzed solely by hydrogen ions.

(b) Electrolyte effect same as sodium chloride.

(c) Activity coefficients vary with ionic concentration in the same manner as those for acetic acid in sodium chloride solution, calculated from data of HARNED & HICKRY.¹¹

much reliance cannot be placed on this value, as Ostwald himself pointed out that the data in dilute solutions are not reliable.

It is evident that, for reactions between non-electrolytes and ions, medium effects of considerable magnitude are to be expected, and these may be specific enough to complicate the kinetic determination of hydrogen ion concentration in concentrated solutions. In non-aqueous solution, it is becoming evident that these specific effects appear at much lower concentrations and that a further understanding of the nature of medium effects is necessary for kinetic studies.

Bibliography

1. OSTWALD, W. 1889. Z. phys. Chem. 3: 369.
2. LEWIS, G. N., & M. RANDALL. 1921. J. Am. Chem. Soc. 43: 1112.
3. BRØNSTED, J. N. 1922. J. Am. Chem. Soc. 44: 938.
4. DEBYE, P., & E. HÜCKEL. 1923. Phys. Z. 24: 185.
5. ONSAGER, L. 1927. Z. phys. Chem. 28: 294.
6. BRØNSTED, J. N. 1922. Z. phys. Chem. 102: 169.
7. CIAPETTA, F. G., & M. KILPATRICK. 1948. J. Am. Chem. Soc. 70: 639.
8. RIESCH, L. C., & M. KILPATRICK. 1935. J. Phys. Chem. 39: 561.
9. OLSON, A. R., & L. K. S. TONG. 1944. J. Am. Chem. Soc. 66: 1555.
10. RIESCH, L. C., & M. KILPATRICK. 1935. J. Phys. Chem. 39: 891.
11. BROCKMAN, F. G., & M. KILPATRICK. 1934. J. Am. Chem. Soc. 56: 1483.

12. KILPATRICK, M., E. F. CHASE, & L. C. RIESCH. 1934. J. Am. Chem. Soc. 56: 2051.
13. KILPATRICK, M. 1940. *Handbuch der Katalyse* II: 258. Julius Springer. Wien.
14. GROVE, C. 1930. J. Am. Chem. Soc. 52: 1404.
15. HARNED, H. S., & B. B. OWEN. 1943. *The Physical Chemistry of Electrolytic Solutions*: 515. Reinhold. New York.
16. SAXTON, B., & T. W. LANGE. 1933. J. Am. Chem. Soc. 55: 3638.
17. MINNICK, L. J., & M. KILPATRICK. 1939. J. Phys. Chem. 43: 259.
18. HANTZSCH, A., & A. WEISSBERGER. 1927. Z. phys. Chem. 125: 251.
19. TAYLOR, H. S. 1913. Medd. K. Vet-Akads. Nobelinstitut 2 (37).
20. HALL, N. F. 1931. Chem. Rev. 8: 191.
21. BAUGHAN, E. C. 1940. Nature. 146: 461.
22. HARNED, H. S., & J. E. HAWKINS. 1928. J. Am. Chem. Soc. 50: 85.
23. HARNED, H. S., & F. C. HICKEY. 1937. J. Am. Chem. Soc. 59: 1284.

THERMAL PROPERTIES OF ELECTROLYTIC SOLUTIONS AND THE DEBYE-HÜCKEL THEORY

BY FRANK T. GUCKER, JR.

Chemical Laboratory, Indiana University, Bloomington, Indiana

Introduction

A great theory presents a picture of Nature which successfully explains the facts for which it was designed. It also successfully predicts new facts, stimulates new experimental methods, and leads to unsuspected generalizations. This is well illustrated by the application of the Debye-Hückel theory of interionic attraction to the thermal properties of electrolytic solutions. Soon after Debye and Hückel¹ gave their theoretical explanation of the activity coefficients of dilute solutions of strong electrolytes, Adams² and Bjerrum³ simultaneously pointed out the application of the Debye-Hückel theory to predicting the heat of dilution of very dilute solutions of strong electrolytes. Bjerrum showed that the sign of the limiting heat of dilution depends upon the dielectric properties of the solvent, and that all salts dissolved in dilute aqueous solutions should *liberate* heat on infinite dilution. The amount of heat depends upon the properties of the solvent and the valence type of the salt. Later, Scatchard⁴ pointed out that the heat of dilution at constant pressure also involves the thermal expansibility of the solvent. He showed that the limiting law at low concentrations for the electrical contribution to the relative apparent molal heat content of a single strong electrolyte, equal to the heat *liberated* when one mole is diluted from a concentration c to concentration zero, is:

$$\Phi L_2 = H_c - H_\infty = -A \left(1 + \frac{d \ln D}{d \ln T} + \frac{1}{3} \alpha T \right) \frac{(\sum \nu_i z_i^2)^{3/2}}{D^{3/2} T^{1/2}} c^{1/2} \quad (1)$$

Here, $A = \sqrt{\pi N^3 e^3 / 1000 k}$, N is Avogadro's number, e the charge of the electron, k is Boltzmann's constant, D the dielectric constant of the solvent, T the absolute temperature, α the coefficient of thermal expansibility of the solvent, ν_i the number of ions of the i^{th} kind, with charge z_i , and c is the molarity. The term $\frac{1}{3} \alpha T$ reduces the limiting slope 7% for water at 25°. $(\sum \nu_i z_i^2)^{3/2}$ is the important *valence factor*.

Gronwall, LaMer and Sandved⁵ extended the Debye-Hückel calculation of the potential at any point in an electrolytic solution, to include higher terms of the exponentials which Debye and Hückel omitted. The complete equation for the electrical contribution to the relative apparent molal heat content of a single electrolyte is given by Scatchard as:

$$\Phi L_2 = H_c - H_\infty$$

$$\text{Here, } H_s = -NkT \sum \nu_i z_i \sum_m \left(\frac{\epsilon^2 z_i^2}{-kTDa} \right)^m \left[\left(1 + \frac{d \ln D}{d \ln T} \right) \frac{X_m(x)}{2} \right. \\ \left. + \alpha T \left(\frac{X_m(r)}{2} - m Y_m(x) \right) - \frac{d \ln a}{d \ln T} \left(X_m(x) - 3m Y_m(x) \right) \right] \quad (2)$$

where a is the mean collision diameter of the ions, $X_m(x)$ and $Y_m(x)$ are complicated functions of z_1/z_2 and x , and:

$$x = \kappa a = \sqrt{\frac{4\pi N \epsilon^2 \sum \nu_i z_i^2 a^3}{1000 kTD}} \quad (3)$$

H_s° is the value of H_s when $x = 0$. If a is independent of temperature, the last term of EQUATION 2 vanishes.

When Bjerrum compared the predictions of the Debye-Hückel theory with the heats of dilution available in the literature, the best series of measurements which he could find were those of Richards and Rowe.⁶ They had studied hydrochloric and nitric acids, and the hydroxides, chlorides, and nitrates of lithium, sodium, potassium, and cesium down to a concentration of 400 moles of water per mole of electrolyte, or 0.14 m . With a sensitivity of 0.001° in the temperature measurements, they could go no lower. At this concentration, some heats of dilution were positive and some negative, and Bjerrum found no agreement with the Debye-Hückel limiting law. FIGURE 1 shows some of these heats of dilution, extrapolated to zero concen-

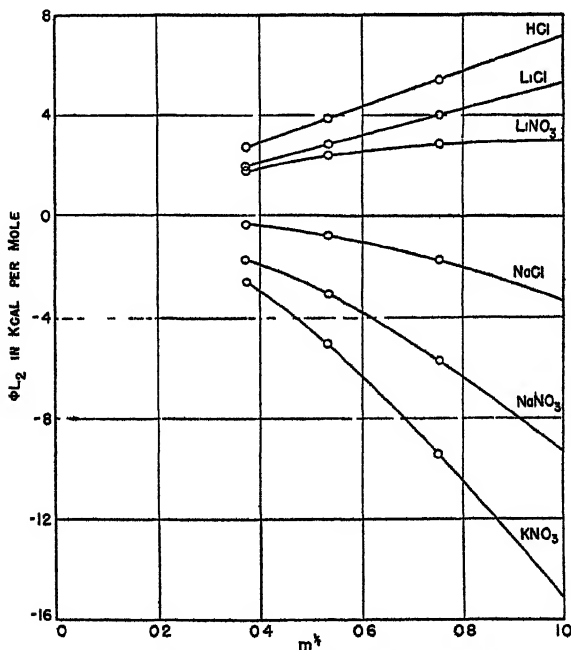


FIGURE 1. Relative apparent molal heat contents of some electrolytes based on extrapolation of the data of RICHARDS & ROWE.

tration and drawn from a common origin so that each curve represents the relative apparent molal heat content of the substance, based on this method of extrapolation.

Microcalorimetry

A test of the predictions of the Debye-Hückel theory required new methods of measuring heats of dilution with greatly increased sensitivity. In 1926, Nernst and Orthmann⁷ used a double calorimeter sensitive to 10 micro-degrees to measure heats of dilution of a number of salts, but could find no agreement with the theoretical predictions. The next year, however, they refined the sensitivity of their previous calorimeter,⁸ and Lange and Messner⁹ described their results with a differential calorimeter of very high sensitivity. Both found that all solutions of representative strong electrolytes liberated heat at great dilution, and that the heats of dilution seemed to approach the limiting slopes predicted by the Debye-Hückel theory.

FIGURE 2 shows the differential microcalorimeter developed by Lange and his co-workers, with which most of the information on heats of dilution of strong electrolytes has been obtained. It consists of an unsilvered Dewar flask, divided into two similar calorimeters, each of about one liter capacity, by a central partition containing a 1,000 to 1,500-junction iron-constantan thermel connected to a sensitive galvanometer. The Dewar flask is immersed in an outer bath, the temperature of which is adjusted to make the system adiabatic. The adiabatic thermel allows a measurement of any temperature difference.

The calorimeters can be filled and emptied through filling tubes, one of which is shown. Each calorimeter contains a pipet, one of which is shown. The solution to be diluted is put into one of the pipets, which also can be filled and emptied from outside. When the temperature difference between the calorimeters has been followed for long enough to establish the temperature trend, the pipet is opened and the solution is diluted. The heat of opening of the pipet may be compensated by opening simultaneously the other pipet, containing water. Most of the heat of dilution is balanced by quantitative electrical heating, and the remainder is read from the temperature difference before and after the experiment. The temperature sensitivity was gradually increased from 5 to about 0.2 microdegree, so that heats of dilution were measured down to 0.0001 *M*.

Another microcalorimeter, developed by Gucker, Pickard, and Planck,¹⁰ is of somewhat different design, as shown in FIGURE 3. Each of the 1-liter twin calorimeters, *C*, *C*, made of tantalum, contains a 60-ml. pipet, *P*, and a stirrer, *S*, only one of which is shown. The calorimeters are suspended inside a submarine jacket, *J*, immersed in a water bath, *B*. The constant for heat conductivity through the 3-cm. air gap to the water bath is 0.003 per minute, compared to 0.004¹¹ to 0.006¹² for a Lange-type calorimeter, using

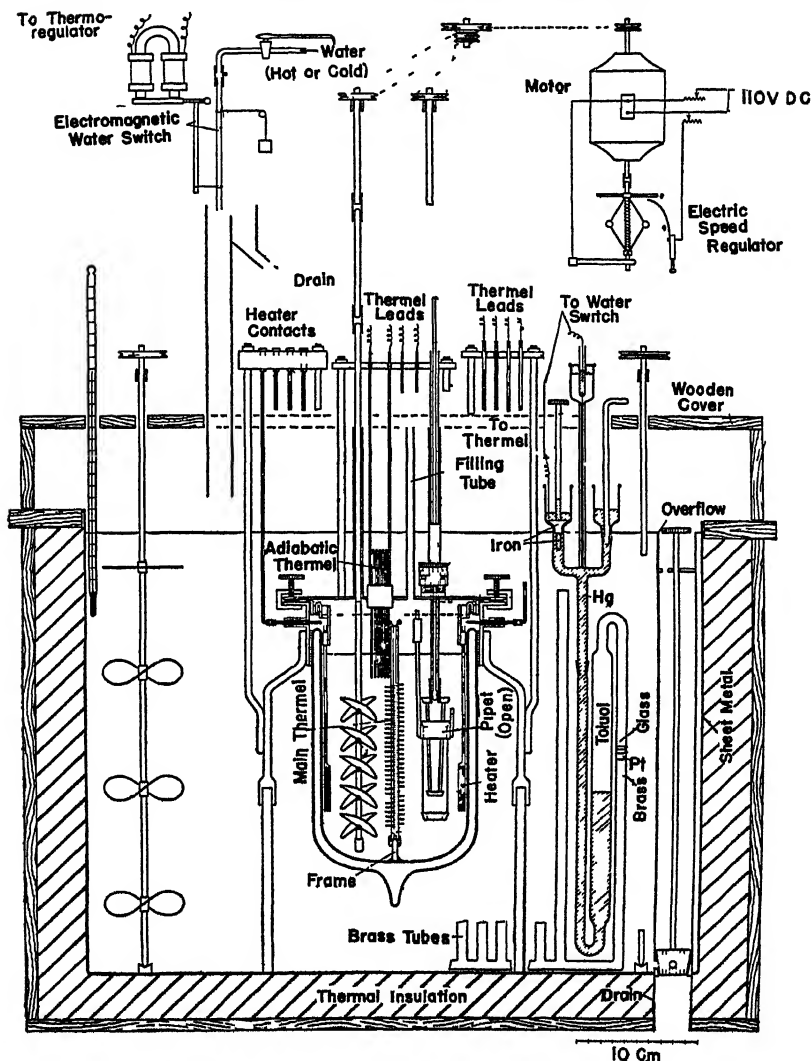


FIGURE 2. Adiabatic differential microcalorimeter of LANGE and co-workers

an *unsilvered* Dewar flask, in which the black-body radiation is the controlling factor. The temperature of the water bath is kept within $\pm 0.0003^\circ$ of that of the calorimeters by an automatic control operated by the adiabatic thermels, *A, A*. The difference in temperature between the two calorimeters is read by means of a 60-junction copper-constantan thermel, *M*, connected to a special low-resistance potentiometer, with a Paschen astatic galvanometer as a null instrument. This gives a sensitivity of 1 mm. per micro-degree. When the pipet *P* is opened by means of the automatic device *F*,

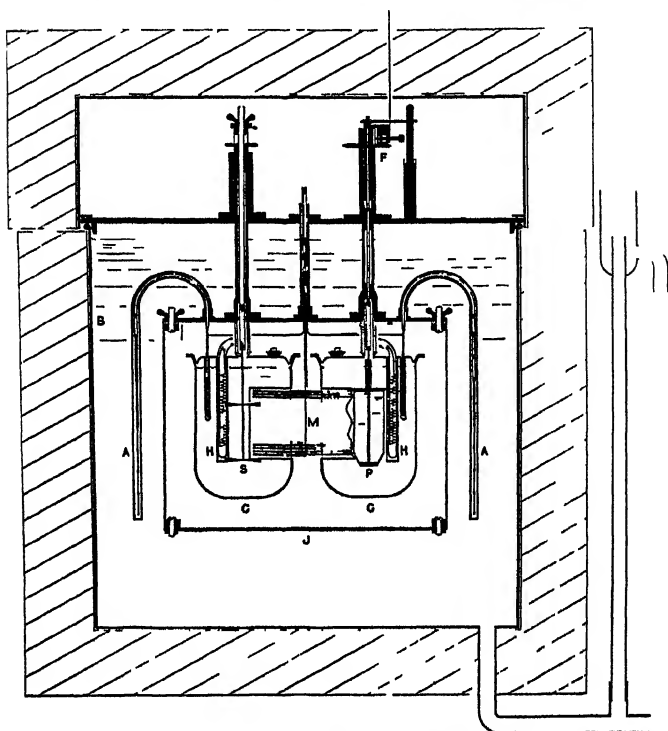


FIGURE 3 Microcalorimeter of GUCLER, PICKARD, & PLANCE. (Courtesy J. Am. Chem. Soc.)

the heat of dilution is balanced by passing an electric current through one of the heaters *H*, *H*. The chief advantage of this type of apparatus is the reduction of the coefficient of thermal conduction between the calorimeters to 0.0011 per minute, which is only about one-twenty-fifth that in the Lange-type apparatus,¹¹ where the two calorimeters are very close together and the thermal junctions are on or near the wall of each calorimeter.

FIGURE 4 illustrates the tremendous increase in calorimetric sensitivity made in a few years, under the impetus of the Debye-Hückel theory. The

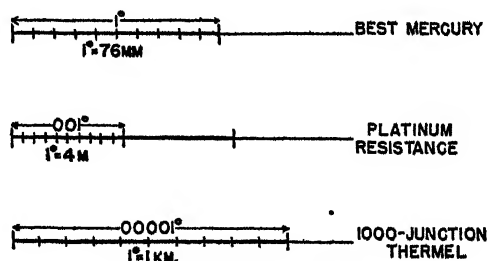


FIGURE 4. Thermometer scales.

thermometric scale was expanded a thousandfold, from a centigrade degree of about three inches for the mercury thermometers of Richards and Rowe, to thirteen feet for a platinum resistance thermometer, and six-tenths of a mile for Lange's thousand-junction thermel and high-sensitivity galvanometer, or the sixty-junction thermel and astatic galvanometer of Gucker, Pickard, and Planck.

Limiting Slopes for Nonelectrolytes and Low-Valence Electrolytes in Water

Before considering the electrolytes, we may examine some measurements, summarized in FIGURE 5, on typical nonelectrolytes, made with our micro-

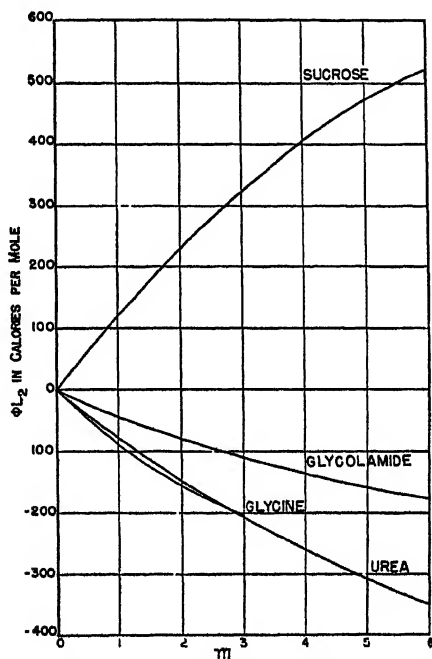


FIGURE 5. Relative apparent molal heat contents of nonelectrolytes

calorimeter.^{10, 13, 14} In every case, the heat of dilution approaches a linear change with the first power of the concentration in dilute solutions, and the relative apparent molal heat content can be expressed as a power series in the concentration with no square root term.

The work of Lange and his co-workers, in Munich, and of Robinson and his co-workers in this country, established the general relationships for the heats of dilution of electrolytes of simple valence type. FIGURES 6, 7, and 8 show the results for 1-1, 1-2, and 2-1 electrolytes, given by Lange and Robinson.¹¹ Their limiting slopes would be reduced by 7% if they had included the term $\frac{1}{2} \alpha T$ in EQUATION 1. In general, the curves show individuality at concentrations as low as 0.0005 *M*, and diverge widely at high

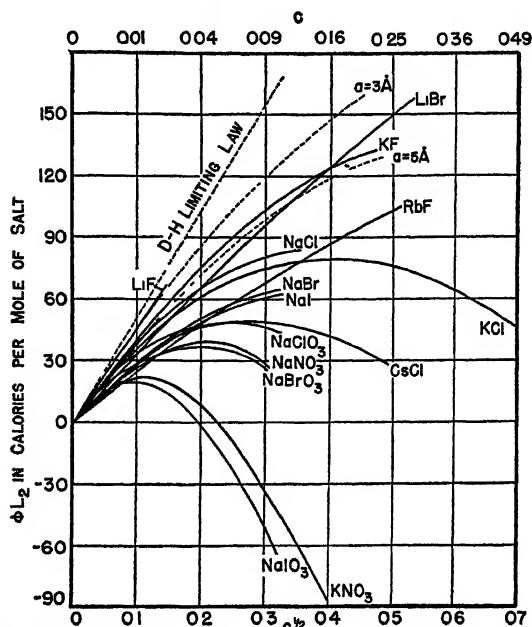


FIGURE 6. Relative apparent molal heat contents of 1-1 salts at 25°C. (Courtesy Chem. Rev., Lange & Robinson)

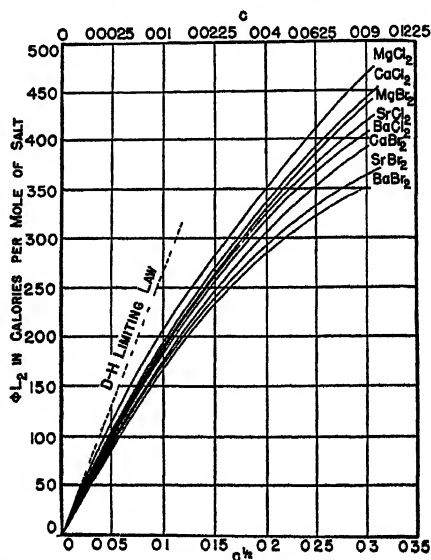


FIG. 7. Relative apparent molal heat contents of 2-1 salts at 25°C (Courtesy Chem. Rev., Lange & Robinson)

concentrations. However, by following the dilution to the lowest possible concentration, a point is reached below which the results are linear with

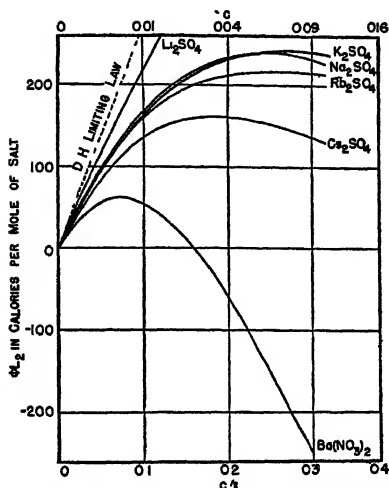


FIGURE 8. Relative apparent molal heat contents of 1-2 salts at 25°C. (Courtesy Chem. Rev., Lange & Robinson)

the square root of the concentration, within experimental accuracy. Lange and Robinson drew a straight line through these points and extrapolated to zero concentration to determine the relative apparent heat contents. The experimental slopes thus determined did not agree quantitatively with the limiting law. For instance, Gulbransen and Robinson,¹² in a careful series of measurements on sodium chloride solutions at 10°, 15°, 20°, and 25° C., found limiting slopes averaging 18 per cent less than the corresponding theoretical limiting slopes calculated from EQUATION 1 by Young and Groenier,¹⁵ using the data of Wyman¹⁶ for the dielectric constant of water.

In order to determine whether or not the data of Gulbransen and Robinson actually were in conflict with the Debye-Hückel limiting law, Young and Groenier¹⁵ introduced a new method of analyzing the data. Instead of plotting the integral heats of dilution, they plotted the values of $\Delta H/\Delta m^{1/2}$ (or $\Delta c^{1/2}$, if volume concentration is used). These correspond to the chords of the heat of dilution curve, or the average value of the slope over the range of measurement, which they designate \bar{P} . If the slope curve can be represented by a quadratic function of $m^{1/2}$ in the dilute region:

$$S = S^\circ + Bm^{1/2} + Cm \quad (4)$$

(which holds up to $m^{1/2} = 0.64$ for NaCl), Young and Groenier showed that the difference between the chord \bar{P} , and the slope P_i at the abscissa $m_i^{1/2}$ corresponding to the midpoint of the chord is:

$$\bar{P}_i - P_i = C\delta_i^2/12 \quad (5)$$

where δ_i is the change in $m^{1/2}$ during the i^{th} dilution. This gives an equation

for the chord:

$$P_i = S^\circ + Bm^{1/2} + C(m_i + \delta_i^2/12) \quad (6)$$

from which the constants of the slope equation can be evaluated by the method of least squares, using the final concentration in each dilution as a weighting factor. Young and Groenier thus determined the limiting slopes from the data of Gulbransen and Robinson, and found values averaging only 2.3 per cent above the theoretical slopes, and agreeing well within the uncertainty of the latter. TABLE 1 shows a comparison of the two methods of

TABLE 1
COMPARISON OF LIMITING SLOPES FOR ΦL_2 vs. $m^{1/2}$

°C	Theory	Gulbransen and Robinson	Δ	Young and Groenier	Δ
10	355	239	-33	356	.3
15	393	340	-13	414	5.2
20	434	370	-15	451	4.0
25	477	418	-12	476	-.2
Av.			-16%		+2.3%

determining the limiting slopes from the same data, together with the percentage difference from the theoretical value, $\Delta \equiv 100 (S^\circ - S_{\text{theory}}^\circ) / S_{\text{theory}}^\circ$. The theoretical slopes differ only slightly from those derived by Harned and Hecker¹⁷ from the same dielectric data. These values, at 5° intervals from 0° to 60° C., are also tabulated on page 125 of the excellent monograph on electrolytic solutions written by Harned and Owen.¹⁸

To illustrate the application of the slope equations in evaluating the relative apparent heat contents of a typical electrolytic solute, we shall consider the case of sodium chloride at 25°. FIGURE 9 shows a plot of the chords of the heats of dilution curve, made by Young and Vogel,¹⁹ from the data of Wüst and Lange,²⁰ Lange and Messner,²¹ and Lipsett, Johnson, and Maass.²² The smooth curve corresponds to the slope which Young and Vogel obtained graphically, before the analytical method of Young and Groenier had been developed. This slope curve was integrated to yield the relative apparent molal heat content curve shown as the solid line in FIGURE 10. The points marked $\Phi H + C$ represent the heats of solution determined by Wüst and Lange, and Lipsett, Johnson, and Maass, plus an additive constant to make the two curves coincide. These points are displaced vertically to correspond with the solid curve, which then gives the extrapolation to zero concentration and allows evaluation of the relative apparent molal heat contents of sodium chloride in solution, and the relative molal heat content of the solid.

Later, Young and Seligmann²³ investigated all the electrolytes for which

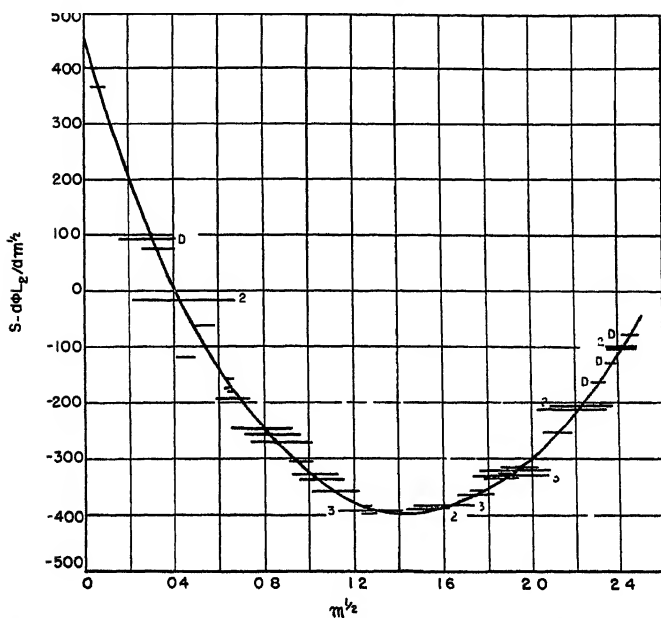


FIGURE 9. Chord-area plot of YOUNG & VOGT for NaCl at 25°. (Numbers refer to overlapping results. Chords determined from heats of dilution are marked D. Courtesy J. Am. Chem. Soc. Young & Vogel)

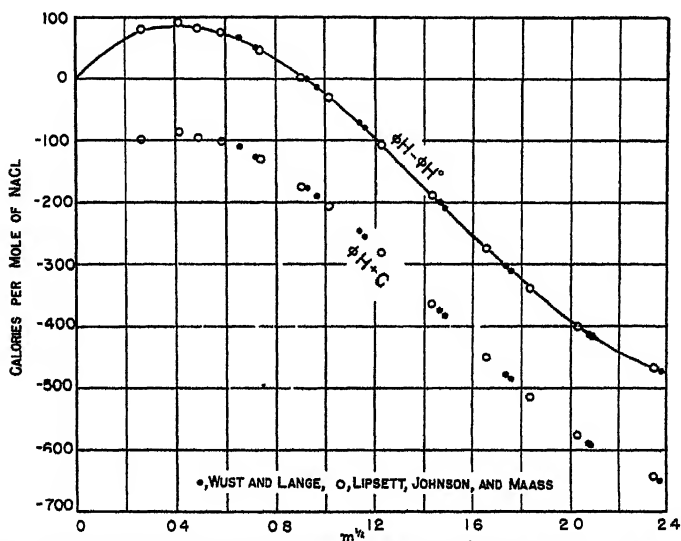


FIGURE 10. Relative apparent molal heat content of NaCl from heats of solution. (Courtesy J. Am. Chem. Soc. Young & Vogel)

sufficient dilution data were available. These were the eleven 1-1 electrolytes and five 1-2 electrolytes shown in FIGURES 11 and 12. The slope equa-

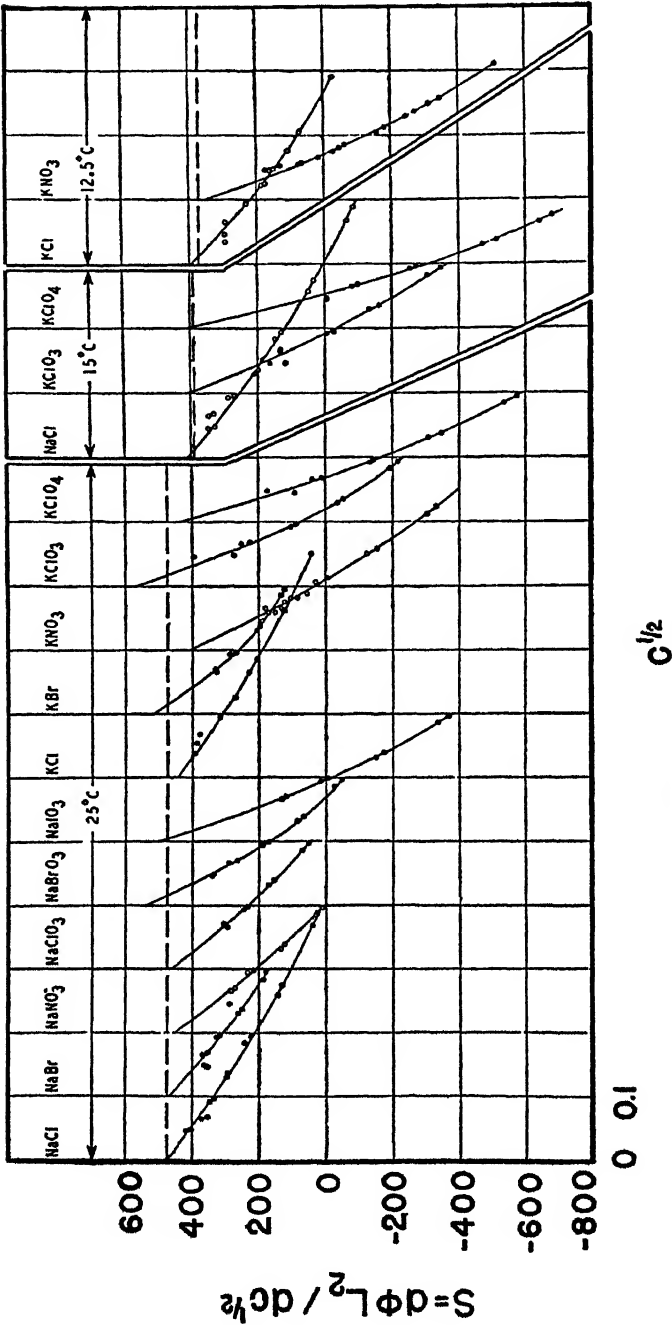


FIGURE 11. Plots of S w. $c^{1/2}$ for 1-1 electrolytes. (Origin of each curve lies one division to the right of that of its predecessor. Broken lines indicate theoretical limit. Courtesy J. Am. Chem. Soc. Young & Seligman)

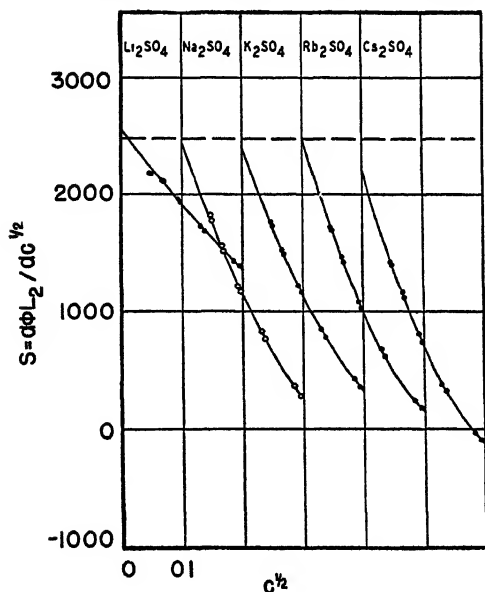


FIGURE 12. Plots of S vs $c^{1/2}$ for 1-2 electrolytes (Origin of each curve lies one division to the right of that of its predecessor. Broken line indicates theoretical limit. Courtesy J. Am. Chem. Soc. Young & Seligman)

tions, determined by the method of least squares, are plotted against $c^{1/2}$ and compared with the theoretical values, shown by the dashed lines. The average agrees with the theoretical within 0.5%, while the deviations average $\pm 6.3\%$. Young and Seligmann concluded that the Debye-Hückel limiting law is completely verified for these salts, within the limit of error of the experimental data. They therefore used the theoretical limiting slope S° and evaluated the constants B and C of EQUATION 4 by means of the method of least squares, to get equations for the relative apparent molal heat content:

$$\Phi L_2 = S^\circ c^{1/2} + \frac{1}{2}Bc + \frac{1}{3}Cc^{3/2} \quad (7)$$

which are valid up to about 0.04 molar.

The chord method of Young and Groenier was modified by Wallace and Robinson²⁴ to make it more readily adaptable to the data obtained by Lange and Robinson and their co-workers, who regularly carried out two successive dilutions with n moles of solute at concentration 1, first with water to a concentration 2, absorbing q_1 calories of heat, and then with the same volume of solution of concentration 2, to a final concentration 3, absorbing q_2 calories. From these data, they calculated and tabulated for the "long-chord" dilutions the heats $\Delta H_{1 \rightarrow 2}$ and $\Delta H_{1 \rightarrow 3}$, from which Young and his co-workers had calculated the long chords, corresponding to a dilution as large as 148-fold. Wallace and Robinson pointed out that the heats of the approximately 2-fold

dilution $\Delta H_{3 \rightarrow 2}$ gave short chords in the dilute region most suitable for the determination of the limiting slope. Robinson and Wallace²⁵ showed that the same value was calculated more easily and accurately from the original heat effects by means of the equation:

$$\Delta H_{3 \rightarrow 2} = (q_1 - q_2)/(2 - \delta)n. \quad (8)$$

Here δ , the ratio of the volumes of the pipet and calorimeter, enters because the pipet full of solution of concentration 2 is removed after the first dilution and refilled with solution of concentration 1 for the second dilution. Substituting the numerical value of δ for their apparatus, these authors used the equation:

$$\Delta H_{3 \rightarrow 2} = 0.5017 (q_1 - q_2)/n. \quad (9)$$

If the value of δ is not stated, it may be calculated from the initial and final volume concentrations, c_1 and c_2 , as:

$$\delta = c_2/(c_1 - c_2). \quad (10)$$

Wallace and Robinson,²⁴ using a linear equation for the slope in the range below 0.005 *m*, obtained a limiting value for sodium sulfate at 25° which differed by only 0.6 per cent from that which Young and Seligmann²³ found by applying a quadratic equation to the data for the long chords up to 0.4 *m*.

Young and Seligmann found no 1-1 or 2-1 electrolytes at variance with the Debye-Hückel limiting law, although the data on the alkaline earth salts were less precise than those shown in FIGURES 8 and 9. Later, Robinson and Wallace²⁵ studied the chlorides, bromides, and iodides of cadmium in aqueous solutions at 15 and 25° and found that even these salts, which are known to be associated in solutions of usual concentration, give results which agree qualitatively at least with the Debye-Hückel theory at the lowest available concentrations. These results show that the Debye-Hückel limiting law correctly describes the change of heat of dilution with valence type of electrolyte and with temperature, for simple salts in aqueous solutions.

Limiting Slopes for 1-1 Electrolytes in Other Solvents

It is interesting, also, to test the theory with other solvents for which the dielectric properties are known; but only a few experimental data can be correlated in this way. Lange and Robinson²⁶ studied the heats of dilution of potassium chloride in 15% aqueous sucrose solution and in 5% aqueous urea solution, and found limiting slopes 3 per cent greater in the former than in water, and 12 per cent less in the latter. No recent precise dielectric constant data are available for the sucrose solutions, but Wyman²⁷ measured the dielectric constant of solutions of urea from 0° to 50° C. and calculated the value of the limiting slope (neglecting the term in α) for a 5% urea solution at 25° as 440. The coefficient of thermal expansibility of urea solutions at 27.5° C. has been calculated by Gucker and Moser²⁸ and given as a

quadratic function of the concentration. Assuming that the other constants are the same at 25°, and calculating the value for water as $\alpha_0 = 2.554 \times 10^{-4}$ from the data of the *International Critical Tables*, we find $\alpha = 2.958 \times 10^{-4}$ for a 5% urea solution at 25°. Interpolating Wyman's results gives 80.96 for the dielectric constant of the solution, whence the α term reduces the limiting slope by 39, giving a theoretical value of 401, compared to 332 determined by Lange and Robinson.

We treated the data of Lange and Robinson by the chord method, determining short chords over three ranges in the dilute region by subtracting successive pairs of results from the same initial concentration to two different final concentrations, and in the more concentrated region by the difference between the values of ΔH for pairs of experiments from different initial concentrations to the same final concentration. The mean error e in measuring any value of ΔH may be considered inversely proportional to the final concentration. When several values of ΔH are combined, E , the mean error in the result, is taken as:²⁰

$$E = \sqrt{\Sigma e^2}. \quad (11)$$

Using the method of Young and Groenier and EQUATION 6, a least square solution of the data was made to determine the best quadratic equation in $c^{1/2}$ for the slope, taking the reciprocal of the mean error as a weighting factor. The resulting equation was:

$$S = 389 - 978 c^{1/2} + 412 c. \quad (12)$$

This gave a limiting slope only 3 per cent below the theoretical. A second least square solution using the theoretical limiting slope was:

$$S = 401 - 1160 c^{1/2} + 925 c. \quad (13)$$

This gives for the relative apparent heat content:

$$\Phi L_2 = 401 c^{1/2} - 580 c + 308 c^{3/2}. \quad (14)$$

In the upper part of FIGURE 13, we have plotted the centers of the chords and the solid line of EQUATION 13, as well as the dashed line of EQUATION 12 and the value of 332, which Lange and Robinson had estimated as the limiting slope. In the lower part of FIGURE 13, we have plotted EQUATION 14 with the experimental data, where $-\Delta H$ for each dilution is added to the value of ΦL_2 calculated for the *dilute* solution. The agreement appears to be within experimental error.

Another example of accurate data in a solvent other than water is the heat of dilution of sodium chloride in ethylene glycol at 25°, measured by Wallace, Mason, and Robinson.³⁰ In this solvent, the theoretical limiting slope they calculated as 1970 for a 1-1 electrolyte. Using the method of Young and Groenier, they found a limiting value of 1580 in a quadratic equation for the slope, with a maximum in the slope at $m^{1/2} = 0.027$. A

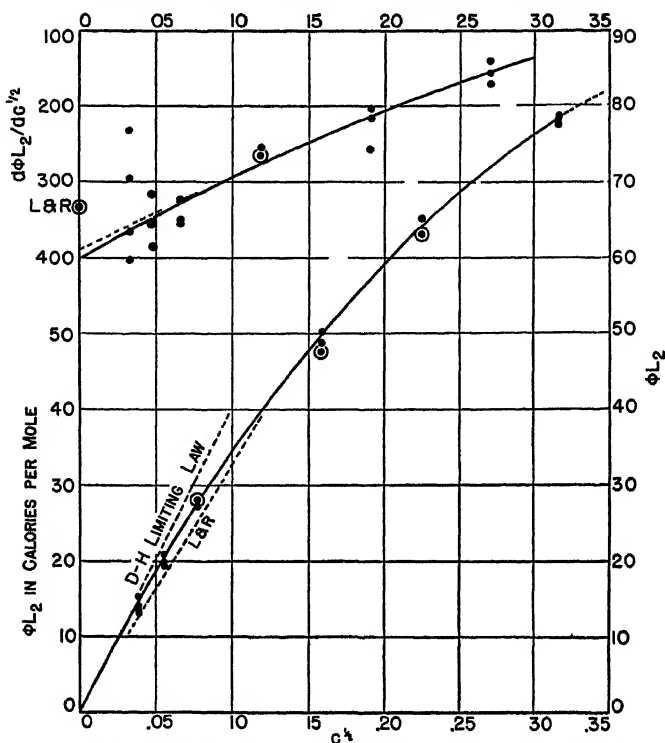


FIGURE 13. Heats of dilution of KCl in 5% aqueous urea solution at 25°.

quadratic equation derived with the theoretical limiting slope fitted the results almost as well as the first equation, and they concluded that the agreement of theory and experiment was within experimental error. The uncertainty was increased by the viscosity of the ethylene glycol, which delayed attainment of thermal equilibrium, and its hygroscopic nature, which caused the absorption of 0.1 to 0.4 per cent moisture. The latter amount, they said, would reduce the limiting slope by 1 per cent. The plot of the results is shown in FIGURE 14, where the full line representing the least-square equation crosses the theoretical slope and approaches it from above.

Limiting Slopes for High-Valent Salts

When Young and Seligmann applied their methods to salts of 2-2 valence type, they discovered that the data available in 1937 did not yield limiting slopes in agreement with theory. In general, the lowest experimental slopes were two or three times the theoretical. However, Young³¹ found some evidence for a maximum in the slope curve of calcium sulfate, and Robinson and Wallace,³² analyzing the data for the sulfates of calcium, magnesium, cadmium, zinc, and copper by their short-chord method, found indications of a

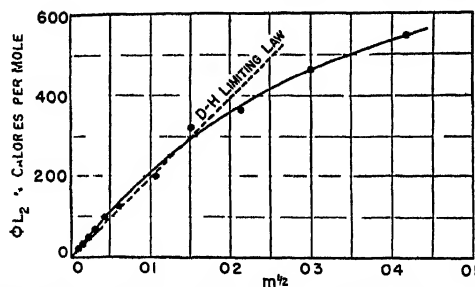


FIGURE 14. Relative apparent molal heat content of NaCl in ethylene glycol at 25°. (Courtesy J. Am. Chem. Soc.)

maximum in the slope between 10^{-3} and $10^{-4} M$. Although the data did not warrant evaluation of the limiting slope, they gave a "strong indication that better agreement with the Debye-Hückel limiting law exists than previously reported."

In the case of lanthanum chloride and sulfate, however, Nathan, Wallace, and Robinson⁴⁴ studied heats of dilution to final concentrations of 6 and $2 \times 10^{-6} M$ and found no evidence of an approach to the limiting law. The slopes at the lowest observable concentrations, as calculated by least squares, were about 0.45 and 17 times the respective limiting slopes required by the theory. It is just such unsymmetrical, high-valent electrolytes for which the Gronwall, LaMer, and Sandved extension of the Debye-Hückel theory predicts large departures from the limiting law at low concentrations, even in a solvent of high dielectric constant like water.

We can easily show that the heats of dilution of all salts may be studied calorimetrically to the same ionic strength, in the region where they obey the limiting law of EQUATION 1. For a dilution from concentration c' to c'' , the limiting law shows that:

$$\Delta H = K(\sum \nu_i z_i^2)^{3/2} ((c'')^{1/2} - (c')^{1/2}) \quad (15)$$

Here, the constant K depends only on the solvent and temperature. If the dilution ratio is fixed, $c' = kc''$ and:

$$\Delta H = K(1 - k^{1/2})(\sum \nu_i z_i^2)^{3/2} (c'')^{1/2} \quad (16)$$

Also, for any dilution:

$$\Delta H = C_p \Delta T = 1000 \Delta T / c'' \quad (17)$$

Combining these equations, we find:

$$\Delta T = K'(\mu'')^{3/2} \quad (18)$$

where $K' = 2^{3/2} K(1 - k^{1/2})/1000$ and $\mu = \frac{1}{2} \sum \nu_i z_i^2$

Thus, the difference in temperature, which limits the calorimetric measurements in dilute solutions, is the same for the same relative dilution of all salts

to the same final ionic strength (μ''), if they obey the limiting law. Unfortunately, this relationship no longer holds when the heat content deviates from the limiting law, and these deviations occur at lower ionic strengths for electrolytes with small, highly-charged ions.

Perhaps the Debye-Hückel second approximation, for ions of finite size, may be useful in extrapolating heats of dilution beyond the limit of experimental investigation, and may explain the data for lanthanum salts. This equation is obtained by setting $m = 1$ in EQUATION 2. The advantage of its use was suggested by Scatchard and Epstein³¹ and by Owen and Brinkley³⁵ at this conference.

Lange and Streek³⁶ have attempted to apply the extended equation of Gronwall, LaMer, and Sandved⁵ and to account for the observed individuality of the curves for heats of dilution of simple electrolytes by different values of the a parameter, corresponding theoretically to the distance of closest approach of the ions. As shown by the dashed lines of FIGURE 6, the relative apparent molal heat content curves fall below the limiting law more rapidly for electrolytes of increasing ionic diameters; hence, the experimental curves should indicate at least the order of the a values. Usually these values parallel the ionic radii in crystals. This is not always the case, but the discrepancies might be due to different degrees of hydration in solution. Certainly, the a values determined from activity coefficient curves should parallel those determined from heats of dilution, if the theory is applied correctly in both cases. Lange and Robinson¹¹ find exactly the reverse to be true in most series they have investigated. Salts which liberate most heat in the dilute region, and hence have small a values judged calorimetrically, show high activity coefficients, and hence have large a values judged from activities. This is true of the alkali and alkaline earth halides. Lange and Robinson concluded that the heats of dilution probably cannot be explained quantitatively in terms of the a values, even for solutions up to 0.01 m .

Above about 0.01 m concentration, the relative apparent molal heat contents of some salts, e.g., sodium iodate and potassium nitrate, fall off rapidly and become negative below 0.1 m , which is hard to explain on the basis of reasonable a values. Also, Lange and Robinson found that curves of salts with one ion in common cross each other in a way which they conclude is inexplicable on the basis of the Debye-Hückel theory. They discuss the influence of incomplete dissociation, and the change of a with temperature and of dielectric constant with concentration, but find no quantitative relationships which account for the heats of dilution beyond the region of very dilute solutions.

Specific Heats and Apparent Molal Heat Capacities

Within the last twenty-five years, several investigators have measured specific heats of solutions to 0.01% or better, using the Joule-Pfaundler twin-calorimeter method. Among them are Richards and Gucker,³⁷ who gave a

long list of historical references, and Randall and Rossini.³⁸ Later, Gucker, Ayres, and Rubin³⁹ introduced several additional refinements. Their apparatus consists of twin 250-cc. gold-platinum alloy calorimeters, having lids of the same material fitted with ground joints, suspended within a submarine jacket like the dilution calorimeters shown in FIGURE 3. A 20-junction thermel connected directly to a sensitive galvanometer measures the temperature between calorimeters to about 10^{-6} degree. The temperature rise in the tare calorimeter, filled with water, is balanced by that in the calorimeter containing solution, as both are heated by electrical resistances connected in series. The chief innovation in this apparatus was the use of heating coils variable in steps of 0.03 per cent, to decrease the heat liberated in the solution, as its heat capacity decreased, so as to balance the temperature rise to within less than 0.1 per cent in a one-degree rise. The arrangement is shown in FIGURE 15. Heaters 1 and 2 are immersed in solution and

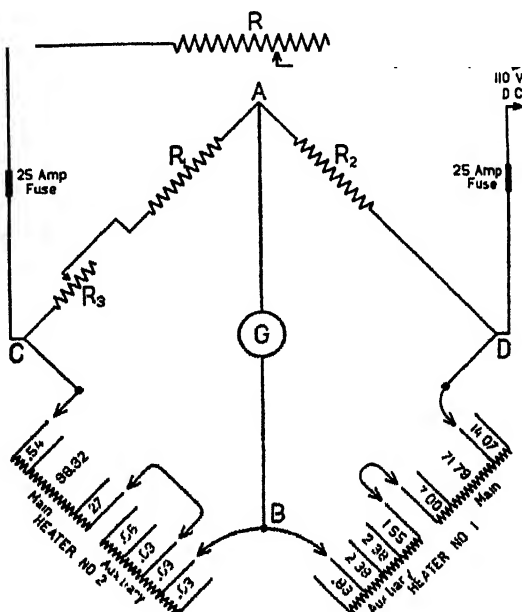


FIGURE 15. Variable heaters and Wheatstone bridge for measuring resistance ratio in determining specific heats. (Courtesy J. Amer. Chem. Soc.)

water respectively. The resistance ratio of the heaters is measured during an experiment by balancing the high-resistance arms of the Wheatstone bridge here shown, by means of R_3 . This apparatus proved flexible and convenient, and has been used from 5° to 85° C.

Randall and Rossini³⁸ calculated the apparent molal heat capacity:

$$\Phi C_{1,2} = \left(\frac{1000}{m} + M_2 \right) s - \frac{1000}{m} \quad (19)$$

where m is the molality of a solute of molecular weight M_2 , and s is the specific heat of the solution relative to water at the same temperature. The apparent molal heat capacity can be determined in calorie units by multiplying by the heat capacity of water at the experimental temperature. Randall and Rossini plotted the apparent molal heat capacity against the square root of the molality and found a linear relationship at low concentrations, which often held over a wide range. This confirmed the conclusion of Randall and Ramage⁴⁰ based on earlier work. Randall and Rossini derived equations connecting the partial molal heat capacities of solute and solvent with the apparent molal heat capacity of the solute. They also showed that the electrical contribution to the relative partial molal heat capacity could be derived from the Debye-Hückel limiting law equation for the partial molal free energy by means of the thermodynamic relationship:

$$\bar{C}_{p_2} - \bar{C}_{p_2}^{\circ} = -T \frac{\partial^2}{\partial T^2} (\bar{F}_2 - \bar{F}_2^{\circ})_P. \quad (20)$$

In carrying out this differentiation, they assumed that the molarity (c) and molality (m) were equal, so that the concentration is independent of temperature. Later, LaMer and Cowperthwaite⁴¹ derived the complete equation:

$$\bar{C}_{p_2} - \bar{C}_{p_2}^{\circ} = \frac{3}{4}A (\sum \nu_i z_i^2)^{3/2} \frac{f(D, V, T)}{(DT)^{3/2}} c^{1/2} \quad (21)$$

where the other symbols have the same meaning as they do in EQUATION 1, and:

$$\begin{aligned} f(D, V, T) = 1 + 2 \frac{d \ln D}{d \ln T} + 5 \left(\frac{d \ln D}{d \ln T} \right)^2 + 2\alpha T \left(\frac{d \ln D}{d \ln T} \right) \\ + \frac{2}{3}\alpha T + \alpha^2 T^2 - 2 \frac{T^2}{D} \cdot \frac{d^2 D}{dT^2} - \frac{2}{3} \frac{T^2}{V} \cdot \frac{d^2 V}{dT^2}. \end{aligned} \quad (22)$$

LaMer and Cowperthwaite's value of $f(D, V, T)$ gives the equation for the limiting slope at 25° C.:

$$\bar{C}_{p_2} - \bar{C}_{p_2}^{\circ} = 4.69(\sum \nu_i z_i^2)^{3/2} c^{1/2} \text{ cal./deg./mole} \quad (23)$$

The coefficient is about 16 per cent lower than the value found by Randall and Rossini, omitting the volume terms. Harned and Hecker¹⁷ also used Wyman's dielectric constants to calculate values of the limiting slopes at 5° intervals from 0° to 60° which are tabulated on page 125 of the monograph by Harned and Owen.¹⁸ The value at 25° C. agrees with that of LaMer and Cowperthwaite. Unfortunately, the theoretical slope is subject to considerable uncertainty in the second temperature derivatives, particularly of the dielectric constant.

The value of the apparent molal heat capacity at infinite dilution, which must depend upon ion-solvent effects, is not predicted by the Debye-Hückel

interionic attraction theory. Zwicky⁴² has attempted to calculate the effect of electrostriction upon the apparent molal heat capacity, but he did not explain the actual deviations among electrolytes of the same valence type, or the observed change with concentration. A fuller discussion of his theory is beyond the scope of this paper, but has been given previously by this author.⁴³

The work of a number of investigators has confirmed the fact that the apparent molal heat capacity varies nearly linearly with the square root of the concentration, over a wide range. Thus, Rossini⁴⁴ tabulated all available accurate data on 1-1 electrolytes, and found that they could be represented by this simple relationship within experimental accuracy, up to concentrations of about 2.5 *m*. The results for lithium chloride at 25° are a typical example,⁴⁵ shown in FIGURE 16. The average deviation from the

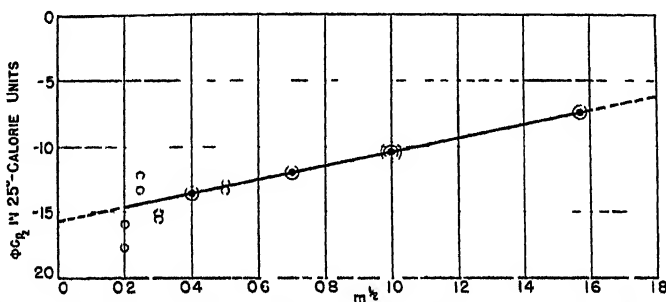


FIGURE 16. Apparent molal heat capacity of LiCl at 25°C. (Courtesy J. Am. Chem. Soc.)

straight line:

$$\Phi C_{p2} = -15.63 + 5.25 m^{1/2} \quad (24)$$

is only $\pm 0.007\%$ in the specific heat.

The best linear representation of the results for a number of 1-1 and 2-1 electrolytes is shown in FIGURE 17. The limiting slopes for the *apparent* molal heat capacities are 2/3 of those calculated from EQUATION 23 for the *partial* molal heat capacities. As pointed out by Randall and Rossini, the experimental lines for salts of the same valence type are not all the same. Although the 2-1 electrolytes show steeper curves than the 1-1 electrolytes, the slopes of the experimental lines are two or three times the theoretical limiting slopes. If the lines start out with the theoretical slope, therefore, they cannot actually be straight up to high concentrations. As pointed out by LaMer and Cowperthwaite,⁴¹ appreciable deviations from the limiting law for salts like zinc sulfate may occur at concentrations as low as 0.0005 *m*, as calculated from the second temperature coefficient of the E.M.F. of cells, and predicted by the Gronwall, LaMer, and Sandved extension of the Debye-Hückel theory.

The values of the apparent molal heat capacity calculated from the best

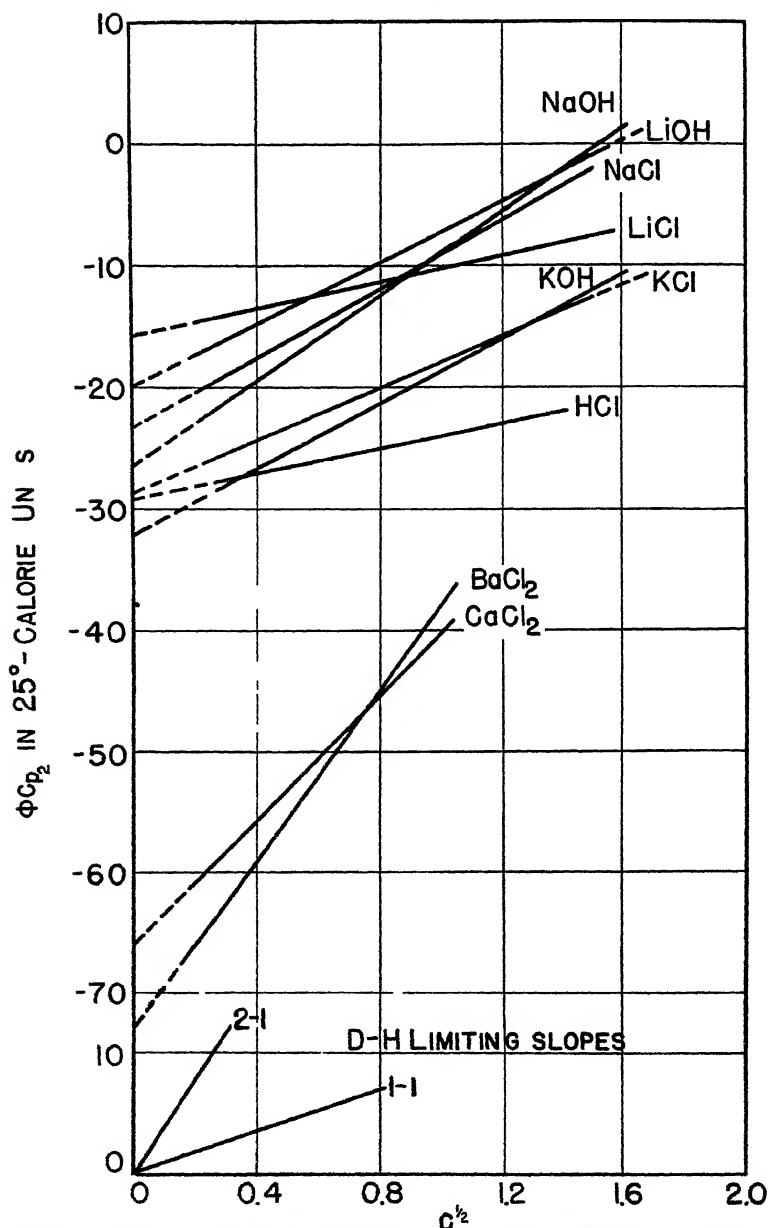
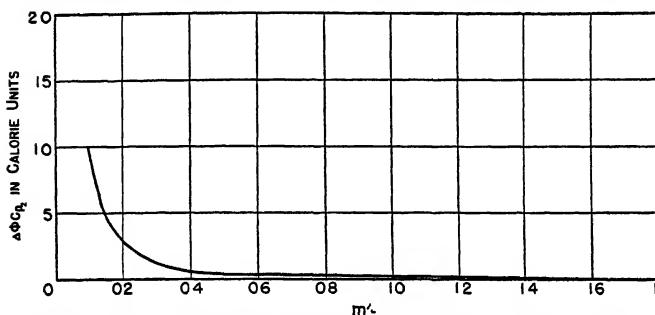


FIGURE 17. Apparent molal heat capacities of some 1-1 and 2-1 electrolytes at 25°C. (Courtesy Chem. Rev.)

available specific heats become increasingly uncertain as the concentration decreases. This is illustrated by the increasing spread of the points in the

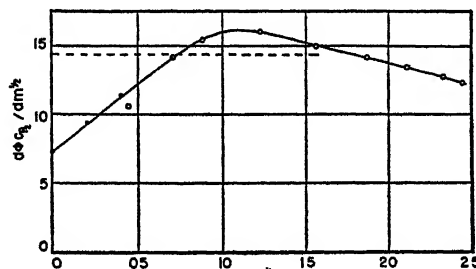
FIGURE 18. Error in ΦC_{p2} for 0.01% error in s . (Courtesy J. Am. Chem. Soc.)

dilute region in FIGURE 16. The error in the apparent molal heat capacity is nearly proportional to the reciprocal of m , as shown in FIGURE 18, so that direct measurements of specific heats much below $0.1\ m$ are useless. Better values of the apparent molal heat capacity can be obtained from the heats of dilution at two temperatures, since these can be measured with about a hundred times the thermometric sensitivity of the specific heats. The relative heat capacities and heat contents are related by the well-known equation:

$$\Phi C_{p2} - \Phi C_{p2}^{\circ} = (\partial \Phi L_2 / \partial T)_P. \quad (25)$$

The uncertainty in the limiting slope for the apparent molal heat capacity is greater than that for the heat content, because it involves the results of two or more extrapolations of the latter quantity.

In their study of sodium chloride solutions, Young and Machin⁴⁶ determined the slope, $d\Phi L_2/dm^{1/2}$, of the apparent relative heat content curves of sodium chloride in aqueous solutions of various concentrations at 12.5° and 25° C. The difference in slope, divided by 12.5° , gives the average value of the slope, $d\Phi C_{p2}/dm^{1/2}$ of the apparent molal heat capacity curve at 18.75° , shown as the solid line in FIGURE 19, taken from their paper. The open circles represent the slopes calculated from the observed chords, applying

FIGURE 19 Slope curve for apparent molal heat capacity of NaCl (Curve, YOUNG & MACHIN, 12.5 – 25° ; dashed line, ROSSINI, 25° . Courtesy J. Am. Chem. Soc.)

the small correction given in EQUATION 5, and the filled circles are derived from the slope equations of Young and Groenier.¹⁵ The slope apparently changes appreciably with concentration, increasing from a value very nearly that predicted by theory, going through a maximum at about 1 *m*, and then decreasing slowly. The dashed line corresponds to the constant slope found by Rossini to fit the data between 18° and 25° C. The deviations are beyond the limit of error of the specific heat measurements. The full line of FIGURE 20 shows the curve for the apparent relative heat capacity obtained by inte-

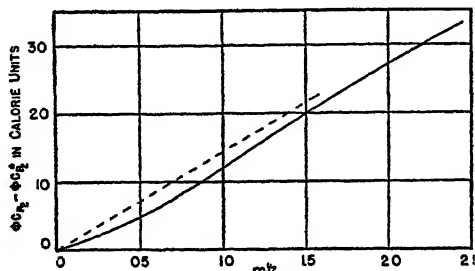


FIGURE 20 Relative apparent molal heat capacity of NaCl. (Curve, YOUNG & MACHIN; dashed line, ROSSINI. Courtesy J. Am. Chem. Soc.)

grating the slope curve of the preceding figure, compared with the dashed curve corresponding to Rossini's linear extrapolation. The two curves are nearly parallel over the range from $m^{1/2} = 0.2$ to $m^{1/2} = 1.6$ of Randall and Rossini's experiments. The difference of about 1.6 calorie units over this range corresponds to the error in the linear extrapolation.

Robinson and his co-workers have calculated the relative apparent molal heat capacities of a number of electrolytes from measurements of heats of dilution. For sodium sulfate, Wallace and Robinson²⁴ found a definite change in the slope below 0.04 *m*, and a limiting value agreeing with theory, within the uncertainty in the calculated results. Robinson and Wallace⁴⁷ gave results for cadmium sulfate at several temperatures, which indicated qualitative agreement with the theoretical limiting slopes. Later²⁸ they found similar results for the chlorides, bromides, and iodides of cadmium. In this case, however, the curves for the apparent molal heat capacities were far from linear functions of the square root of the molality at higher concentrations. They attributed these complications to association of the ions, for which there has been ample evidence from other anomalies of these substances.

Summary

The Debye-Hückel interionic attraction theory of electrolytes has influenced profoundly the development of thermochemistry in the last twenty-five years. The sensitivity of measuring heats of dilution has been increased more than a thousand fold. The microcalorimetric measurements, inter-

preted by correct mathematical analysis, have provided one of the most striking verifications of the Debye-Hückel theory in the limiting slopes for the heats of dilution of strong 1-1, 2-1, and 1-2 electrolytes at different temperatures.

Methods of measuring specific heats of solutions have been improved greatly during the same period. The apparent molal heat capacities of simple strong electrolytes determined directly from specific heats or from the temperature coefficients of heats of dilution, also are found to approach the limiting law predicted by the Debye-Hückel theory.

The individuality of heats of dilution of simple strong electrolytes at all but the lowest concentrations, and of high-valence electrolytes even at extreme dilution, awaits a satisfactory explanation. The same is true of the nearly-linear relation between the apparent molal heat capacity and the square root of the molality, and the actual value of the apparent molal heat capacity at infinite dilution.

Bibliography

1. DEBYE, P., & E. HÜCKEL. 1923. *Phys. Z.* **24**: 185.
2. ADAMS, E. Q. 1926. *J. Am. Chem. Soc.* **48**: 621.
3. BJERRUM, N. 1926. *Z. phys. Chem.* **119**: 145.
4. SCATCHARD, G. 1931. *J. Am. Chem. Soc.* **53**: 2037.
5. GRONWALL, T. H., V. K. LAMER, & K. SANDVED. 1928. *Phys. Z.* **29**: 358.
6. RICHARDS, T. W., & A. W. ROWE. 1920. *J. Am. Chem. Soc.* **42**: 1621; 1921. **43**: 770.
7. NERNST, W., & W. ORTHMANN. 1926. *Sitz. preuss. Akad. Wiss.* **51**.
8. NERNST, W., & W. ORTHMANN. 1927. *Sitz. preuss. Akad. Wiss.* **136**.
9. LANGE, E. W., & G. MESSNER. 1927. *Z. Elektrochem. angew. phys. Chem.* **33**: 431.
10. GUCKER, FRANK T., JR., H. B. PICKARD, & R. W. PLANCK. 1939. *J. Am. Chem. Soc.* **61**: 459.
11. LANGE, E., & A. L. ROBINSON. 1931. *Chem. Rev.* **9**: 89.
12. GULBRANSEN, E. A., & A. L. ROBINSON. 1934. *J. Am. Chem. Soc.* **56**: 2637.
13. GUCKER, F. T., JR., & H. B. PICKARD. 1940. *J. Am. Chem. Soc.* **62**: 1464.
14. GUCKER, F. T., JR., H. B. PICKARD, & W. L. FORD. 1940. *J. Am. Chem. Soc.* **62**: 2698.
15. YOUNG, T. F., & W. L. GROENIER. 1936. *J. Am. Chem. Soc.* **58**: 187.
16. WYMAN, J. 1930. *Phys. Rev.* **35**: 623.
17. HARNED, H. S., & J. C. HECKER. 1933. *J. Am. Chem. Soc.* **55**: 4838.
18. HARNED, H. S., & B. B. OWEN. 1943. *The Physical Chemistry of Electrolytic Solutions*. Reinhold. New York.
19. YOUNG, T. F., & O. G. VOGEL. 1932. *J. Am. Chem. Soc.* **54**: 3030.
20. WÜST, J., & E. LANGE. 1925. *Z. phys. Chem.* **116**: 161.
21. LANGE, E., & G. MESSNER. 1927. *Naturwissenschaften* **15**: 521; 1927. *Z. Elektrochem.* **33**: 431.
22. LIPSETT, S. G., F. M. G. JOHNSON, & O. MAASS. 1927. *J. Am. Chem. Soc.* **49**: 925.
23. YOUNG, T. F., & P. SELIGMANN. 1938. *J. Am. Chem. Soc.* **60**: 2379.
24. WALLACE, W. E., & A. L. ROBINSON. 1941. *J. Am. Chem. Soc.* **63**: 958.
25. ROBINSON, A. L., & W. E. WALLACE. 1942. *Chem. Rev.* **30**: 195.
26. LANGE, E., & A. L. ROBINSON. 1930. *J. Am. Chem. Soc.* **52**: 4218.
27. WYMAN, J. 1933. *J. Am. Chem. Soc.* **55**: 4116.
28. GUCKER, F. T., JR., & C. E. MOSER. 1939. *J. Am. Chem. Soc.* **61**: 1558.
29. MELLOR, J. W. 1946. *Higher Mathematics for Students of Chemistry and Physics*: 527. Dover Publications. New York.
30. WALLACE, W. E., L. S. MASON, & A. L. ROBINSON. 1944. *J. Am. Chem. Soc.* **66**: 362.
31. YOUNG, T. F. 1937. *Sci.* **85**: 48.
32. ROBINSON, A. L., & W. E. WALLACE. 1941. *J. Am. Chem. Soc.* **63**: 1582.
33. NATHAN, C. C., W. E. WALLACE, & A. L. ROBINSON. 1943. *J. Am. Chem. Soc.* **65**: 790.

34. SCATCHARD, G., & L. F. EPSTEIN. 1942. *Chem. Rev.* **30**: 211.
35. OWEN, B. B., & S. R. BRINKLEY, JR. *Ann. N. Y. Acad. Sci.* **50** (4): 753.
36. LANGE, E., & H. STREEK. 1931. *Naturwissenschaften* **19**: 359.
37. RICHARDS, T. W., & FRANK T. GUCKER, JR. 1925. *J. Am. Chem. Soc.* **47**: 1876.
38. RANDALL, M., & F. D. ROSSINI. 1929. *J. Am. Chem. Soc.* **51**: 323.
39. GUCKER, FRANK T., JR., F. D. AYRES, & T. R. RUBIN. 1936. *J. Am. Chem. Soc.* **58**: 2118.
40. RANDALL, M., & W. D. RAMAGE. 1927. *J. Am. Chem. Soc.* **49**: 93.
41. LAMER, V. K., & I. A. COWPERTHWAIT. 1933. *J. Am. Chem. Soc.* **55**: 1004.
42. ZWICKY, F. 1925. *Physik. Z.* **26**: 664; 1926. *Ibid.* **27**: 271; 1926. *Proc. Nat. Acad. Sci.* **12**: 86.
43. GUCKER, F. T., JR. 1933. *Chem. Rev.* **13**: 111.
44. ROSSINI, F. D. 1931. *J. Res. N. B. S.* **7**: 47.
45. GUCKER, F. T., JR., & K. H. SCHMINKE. 1932. *J. Am. Chem. Soc.* **54**: 1358.
46. YOUNG, T. F., & J. S. MACHIN. 1936. *J. Am. Chem. Soc.* **58**: 2254.
47. ROBINSON, A. L., & W. E. WALLACE. 1941. *J. Am. Chem. Soc.* **63**: 1582.

INTERACTIONS BETWEEN VAPORS AND HIGH POLYMERS*

BY MALCOLM DOLE

Department of Chemistry, Northwestern University, Evanston, Illinois

Statistical Theories of Sorption Phenomena

When a solid absorbs or adsorbs vapors, the phenomenon may be treated from the standpoint of mixing a liquid with the solid, or of mixing two liquids, one of them being in an elastomer above its melting point, or the phenomena may be treated from the point of view of the sorption of the vapor on a number of localized sites within the solid. If the entropy of the solid itself is not changed by the sorption, as it would be, of course, if the solid partially dissolved in the sorbed vapors, then it seems logical to calculate the free energy of mixing solely from the standpoint of sorption of the vapors on a number of specific, localized sites. Other necessary assumptions are that the sorbed gas is immobile, that the number of sorption sites, N_s , does not change with the amount of material sorbed (probably not strictly true), that a thermodynamic equilibrium exists between the sorbed molecules and the molecules in the vapor state, and that there is no interaction between the sorbed molecules.

Making use of these assumptions and the statistical methods of Fowler and Guggenheim,¹ Cassie,² and Hill,³ the author⁴ derived the following general sorption isotherm

$$\frac{N}{N_s} = a \cdot \frac{\Phi'}{\Phi} \quad (1)$$

where N is the number of moles of vapor sorbed per unit of adsorbent, N_s is the number of moles of sorption sites per unit of adsorbent, a is the relative vapor pressure or activity of the vapor being sorbed, Φ' the first derivative of Φ with respect to a , and Φ is given by the series

$$\Phi = 1 + c_1 a + c_1 \cdot c_2 \cdot a^2 + c_1 \cdot c_2 \cdot c_3 \cdot a^3 + \dots \quad (2)$$

The constants c_1 , c_2 , etc., are the ratio of the internal partition functions of the sorbed molecules in the first, second, etc., layers to that in the pure liquid. If a linear harmonic oscillator model is assumed both for the sorbed and liquid states, and if it is assumed that the sorbed molecules no longer have the communal entropy of the pure liquid (this assumption will become less true, the greater the sorption), then the constant c_1 , for example, will be given by the expression

$$c_1 = \left(\frac{\nu_L}{\nu}\right)^3 \frac{Q_1}{Q} \cdot \frac{1}{a} \cdot e^{-(\psi_1 - \psi_L)/RT}, \quad (3)$$

* This work is part of a general program of study of the thermodynamics of high polymers, for which grants have been gratefully received from E. I. DuPont de Nemours and Company, the Research Corporation (Frederick Gardner Cottrell Grant), the Richardson Company, and the Visking Corporation.

where ν_L and ν_1 represent the vibration frequencies of the molecule as a whole for the pure liquid and sorbed states, respectively. The symbol Q stands for internal molecular vibrational and rotational contributions to the partition function; $1/e$ represents the communal entropy contribution, while $\psi_1 - \psi_L$ is the heat of sorption less the heat of condensation at the absolute zero, the sign of $\psi_1 - \psi_L$ being negative for a large evolution of heat on sorption of the vapors.

At this point, it should be noted that, from the BET⁵ constant c_1 , the heat of sorption in the first layer has often been calculated, the equation

$$\overline{\Delta H_1} = -RT \ln c_1 \quad (4)$$

being used. The statistical analysis shows that such a treatment is not correct; a better approximation is to assume that all the terms of EQUATION 3 are independent of the temperature except the exponential. If the change of c with temperature is known, the quantity $\psi_1 - \psi_L$ could then be calculated from the derivative

$$\frac{\partial \ln c_1}{\partial T} = \frac{\psi_1 - \psi_L}{RT^2}. \quad (5)$$

However, $\psi_1 - \psi_L$ is still not identical with ΔH_1 , unless the variation in $\psi_1 - \psi_L$ from the absolute zero to the temperature in question is neglected.

In the case of the sorption of water by collagen, the initial net heat of sorption (sometimes called the heat of swelling) is equal to -6600 cal./mole.⁶ Bull⁷ gives the following values of c_1 for collagen:

$$c_1 \text{ at } 25^\circ \text{ C.} = 22.4$$

$$c_1 \text{ at } 40^\circ \text{ C.} = 13.1$$

from which $\psi_1 - \psi_L$ can be calculated from EQUATION 5 (after integrating between 25° and 40° C.) to be -6636 cal./mole, a remarkable agreement. However, in the case of other substances such as drawn nylon, there is no agreement at all (the sign even comes out wrong!). Obviously, at the present time it is not possible to calculate the heat of sorption from the constant c_1 with any degree of assurance that the answer is even qualitatively correct.

If the c_1 constants of EQUATION 2 are all unity (the simplest assumption), and if the number of layers is infinite, EQUATION 1 reduces to

$$\frac{a}{N} = \frac{1}{N_s} - \frac{a}{N_s}, \quad (6)$$

which can be considered as Raoult's law, taking the solution to be that of N molecules of the solvent distributed among N_s sites.

If the c_1 constants of EQUATION 2 are equal to some constant k less than unity, EQUATION 1 reduces to (for an infinite number of layers)

$$\frac{a}{N} = \frac{1}{kN_s} - \frac{a}{N_s}. \quad (7)$$

If the c_1 constant for the first layer is finite and the constants for the other layers all zero (no sorption beyond the first layer), Langmuir's equation results:

$$\frac{a}{N} = \frac{1}{c_1 N_s} + \frac{a}{N_s} \quad (8)$$

If the c_1 constants are unity for all layers except the first, the well-known BET⁵ equation for multi-layer sorption is obtained:

$$\frac{1}{N(1-a)} = \frac{1}{c_1 N_s} + \frac{c_1 - 1}{c_1} \cdot \frac{a}{N_s} \quad (9)$$

If the c_1 constants decrease harmonically, that is to say, if

$$c_2 = 1/2c_1, \quad c_3 = 1/3c_1, \text{ etc.},$$

$$\frac{a}{N} = \frac{1}{c_1 N_s} \quad (10)$$

in other words, a linear sorption isotherm is obtained. It should be noted that this is true even for the case of multi-layer sorption.

In the application of the above equations to the experimental data (see "*Applications*"), we shall find it convenient to plot the function a/N against a , inasmuch as if the data follow the Langmuir equation, a straight line will be obtained having a positive slope equal to $1/N_s$; if the data follow Raoult's law, a straight line will again be observed, but with negative slope equal to $1/N_s$; if the modified Raoult's law, EQUATION 7, is followed, a/N will extrapolate to a finite value at a equal to unity; and, finally, if a/N is constant, the validity of a linear sorption isotherm is demonstrated.

Barrer⁸ has suggested that an adsorption isotherm be constructed by combining an equation for the entropy of mixing derived by Miller⁹ with a heat term obtaining in the simplest case

$$\left[1 - \left(1 - \frac{1}{n} \right) \frac{2}{z} v_r \right]^{z/2} \cdot n_1 = \left(\frac{kT}{2\pi m} \right)^{3/2} \frac{1}{v^3} e^{\psi/RT} \quad (11)$$

where v_0 is the volume fraction of the solvent, v_r the volume fraction of the solid, z a coordination number, n the number of segments per polymer molecule, n_1 the number of molecules per ml. in the gas, and the other quantities have their usual meanings. We have attempted to apply this equation to the sorption of water by polyvinyl alcohol without success.

White and Eyring¹⁰ have recently reviewed the subject of the sorption of water by "swelling high polymeric materials." In essence, their treatment consists in adding a third variable constant to the BET equation to allow for the entropy of swelling. Inasmuch as it is not possible to estimate the magnitude of this term for most of the polymers to be discussed, we shall not consider further the sorption isotherm of White and Eyring.

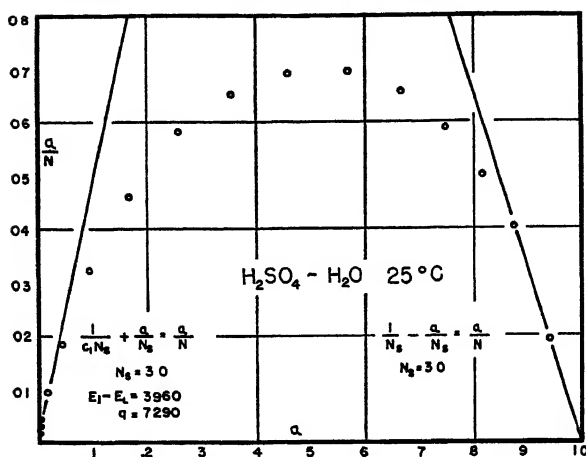


FIGURE 1.

Application to Some Typical Systems

Water—Sulfuric Acid. FIGURE 1 illustrates the sorption of water by sulfuric acid plotted according to the a/N function. Here, N is the concentration of water in the sulfuric acid calculated in terms of moles of water per mole of sulfuric acid and a is the activity of the water. At low concentrations of water, the Langmuir equation (8) is approximated with N_s equal to 3 moles of sorption sites per mole of sulfuric acid. At high water concentrations, the linear function demanded by Raoult's law, EQUATION 6, is also approximated with N_s again equal to 3. If the data were plotted in accordance with the linear BET function $a/(N(1 - a))$, fair agreement with a straight line would be obtained at the lowest concentrations with N_s equal to 3.8.

Water—Collagen. In FIGURE 2, we have plotted the a/N function for the water sorption by a typical protein, collagen. In this case, N is the number of moles of water per 100 g. of collagen, the data being those of Bull.⁷ From the low activity end of the curve, N_s is calculated to be 0.56 moles of sorption

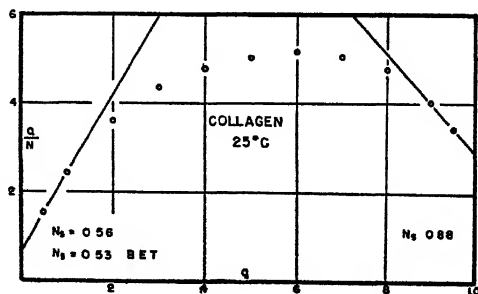


FIGURE 2.

sites per 100 g. of protein, while the BET function, FIGURE 3, extrapolates to N_s equal to 0.53 in satisfactory agreement with the calculation based on

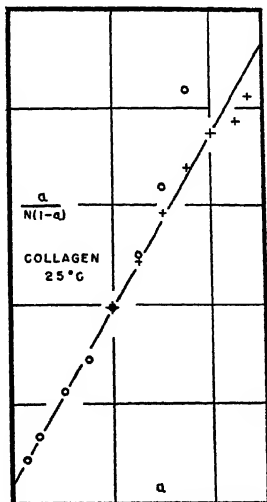


FIGURE 3.

Langmuir's equation. As the water sorption by collagen is not infinite, a/N does not go to zero at a equal to unity but extrapolates to a value slightly less than 0.3. It would be interesting to have water sorption values at water activities between 0.95 and unity in order to determine, more accurately than is done in FIGURE 2, the slope of the a/N function in the region where the modified Raoult's law, EQUATION 7, becomes more valid. If the slope were known accurately here, it might be possible to answer, at least tentatively, the question with respect to the constancy of N_s over the whole sorption range.

Because of the limited sorption of water by collagen, it is reasonable to apply the limited layer BET equation, which, put into linear form, is

$$\frac{a}{N(1-a)} [1 - (n+1)a^n + na^{n+1}] + \frac{a^{n+1}}{N_s} = \frac{1}{c_1 N_s} + \left(\frac{1}{N_s} - \frac{1}{c_1 N_s} \right) a. \quad (12)$$

The crosses of FIGURE 3 are values calculated from the left side of EQUATION 12, which should be proportional to a , taking n equal to 7 and $1/N_s$ equal to 1.86. Satisfactory agreement with a linear relation is seen to exist up to $a = 0.8$. Furthermore, if a/N , FIGURE 2, is extrapolated to a equal to unity to obtain the saturation value of N , 3.5 moles of water per 100 g. of collagen results. Dividing 3.5 by 0.53, we obtain 6.6, which agrees with 7 within the uncertainties of the extrapolation to a equal to unity.

Another test of the validity of applying the BET equation to collagen is

to consider the heat involved in the sorption. According to the chief postulate of the BET theory, the entire heat of sorption should be that of the sorption on the first layer. Dole and McLaren⁶ have estimated the partial molal net heat of sorption (sometimes called the heat of swelling) for the first water to be sorbed by collagen to be -6600 cal./mole, and the total heat of sorption up to saturation to be -4140 cal./mole. Multiplying -6600 by 0.53 , the number of moles of water sorbed in the first layer per 100 g. of collagen, we obtain -3500 . Thus, about 85 per cent of the total heat of sorption is accounted for by the sorption in the first layer.

Water—Polyvinyl Alcohol. Inasmuch as polyvinyl alcohol dissolves in water, one should expect the water sorption phenomenon here to be different from that of collagen, as, indeed, turns out to be the case. A plot of a/N versus a reveals (FIGURE 4) that a/N is constant up to 30 per cent relative

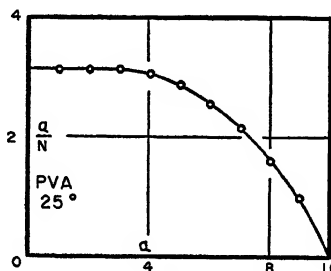


FIGURE 4

humidity and then gradually declines to zero. Over no range of values is the BET equation (9) valid (FIGURE 5). Introducing the parameter n into

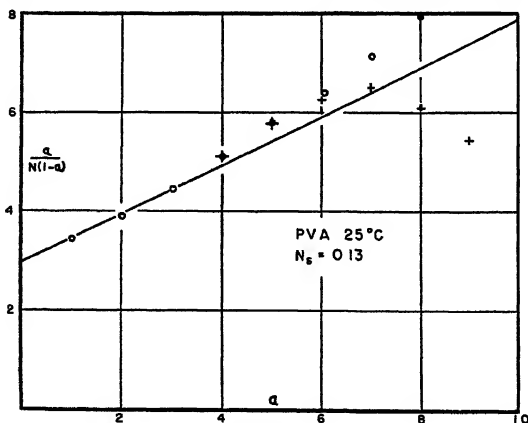


FIGURE 5.

the limited layer BET equation (12) fails to bring about any improvement between the data and the BET theory (crosses calculated for $n = 10$). The polyvinyl alcohol data are those of Hauser and McLaren.¹¹

Water—Nylon. Both drawn and undrawn nylon sorb water in much the same manner as collagen, but the $a/N - a$ curves (FIGURE 6) seem to be

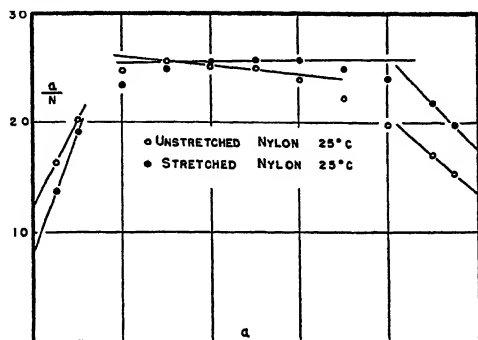


FIGURE 6.

somewhat flatter at intermediate a values. Again the sorption is limited, with the slopes of the $a/N - a$ curves smaller at high values of a than at low. In FIGURE 6, the N values are calculated in terms of moles per mole of carbonyl group in the nylon. N_s , calculated from the slopes of the curve turns out to be approximately 0.1, which means that only one out of every ten polar groups—peptide bonds—in the nylon are able to sorb water. This seems to be a general phenomenon among polymers of this type. Probably the high crystallinity of the nylon accounts for the relatively small number of sorption sites. We hope to discuss this problem further in a later publication, in which data for the sorption of water by a number of isobutylated 6-10 nylons will be presented.

Sorption of Vapors by Rubber Hydrocarbons

Acetone—Rubber. FIGURE 7 illustrates data for the sorption of acetone

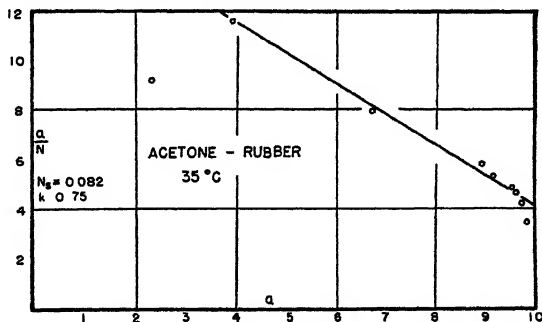


FIGURE 7

by rubber, in which N is calculated as the moles of acetone sorbed per mole of isoprene unit in the rubber plotted according to the Langmuir-Raoult's Law relation, a/N , as a linear function of a . The data, which are those of

Lens,¹² are not extensive enough at low activities of the acetone to enable the curve to be drawn in this concentration range. The one point at a equal to 0.23 suggests that the complete curve in this case would pass through a maximum, similarly to the curves for nylon, collagen, and many other water sorption systems. An unusual feature of the acetone-rubber curve is the long linear portion at high values of the activity of the acetone. It appears that such behavior is characteristic of the sorption of organic vapors by rubber.

Toluene—Rubber. The data of Meyer, Wolff, and Boissonnas¹³ for the sorption of toluene by rubber (FIGURE 8) also follow the linear relation be-

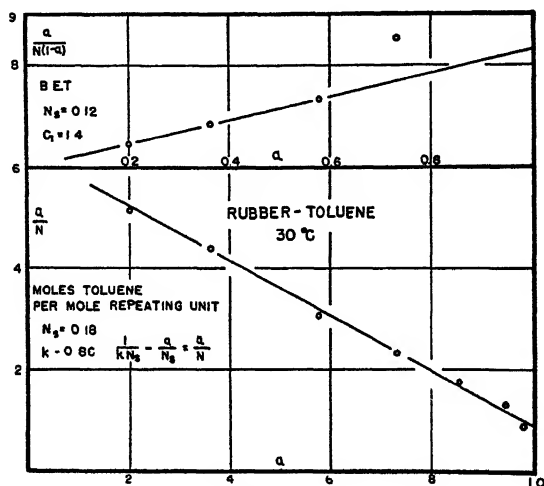


FIGURE 8.

tween a/N and a over a wide range of activities from $a = 0.2$ to $a = 0.95$. Above this activity, the data deviate from this relationship, as is shown by the fact that rubber hydrocarbons dissolve in toluene (a/N going to zero at a equal to unity). The data for the rubber-toluene system can best be understood in relation to the rubber-benzene data.

Benzene—Rubber. Gee and Treloar¹⁴ have made an extensive study of the benzene-rubber system, from which they have calculated and tabulated the partial molal free energy, heat, and entropy of the benzene at different weight fractions of the benzene at 25° C. From these data, it is possible to calculate a/N and a , which we have done and plotted in FIGURE 9. The sorption of benzene by rubber can then be seen to agree with the modified Raoult's law, EQUATION 7, over a very extensive activity range, from zero activity up to 0.8 activity of the benzene. In this treatment, we have calculated N in terms of moles of benzene per mole of isoprene unit in the rubber.¹³

From the slope of the straight line and from this intercept at a equal to unity, it is possible to calculate the constants of the equation, N_s and k . In

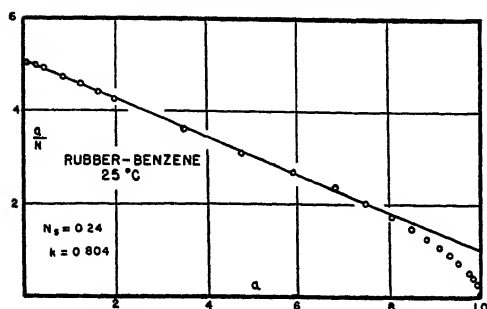


FIGURE 9

TABLE 1, we have collected values of these constants for the sorption of acetone, toluene, and benzene by rubber.

It is very interesting to note that the constants k , for the different vapors, are much more alike than are the N_s values. In the case of benzene, a N_s value of 0.24 means that, on the average, there is only one sorption site for a benzene molecule per four isoprene units in the rubber. In the case of toluene, five isoprene units are required to accommodate one toluene molecule according to this theoretical method of analysis. It is not difficult to conceive of each benzene or toluene molecule as surrounded by four or five repeating units in the rubber. If the data for toluene are plotted according to the BET function, a rough estimate of N_s can be made from the data at the three lowest concentrations of toluene in the rubber (see FIGURE 8); the value 0.12 for N_s results, which is the same order of magnitude as that given in TABLE 1.

TABLE 1
CONSTANTS OF EQUATION 7

Vapor	N_s	k
Acetone	0.082	0.75
Toluene	0.18	0.86
Benzene	0.24	0.80

In order to calculate the partial molal entropy of benzene in the rubber-benzene mixtures, it is necessary to know the relative partial molal heat content, $\overline{\Delta H}_1$. Gee and Treloar¹⁴ listed exact values in terms of calories per gram of benzene as a function of the weight fraction of benzene in the rubber, from which we have calculated $\overline{\Delta H}_1$. On plotting these data of Gee and Treloar as a function of the activity of the benzene* (FIGURE 10, solid line), we were surprised to find that $\overline{\Delta H}_1$ decreased linearly with a . However, the more recent data of Gee and Orr¹⁵ do not follow this simple relationship, their

* Throughout this paper, we have assumed that the partial vapor pressure ratio, p/p_0 , is equal to the activity of the vapors concerned.

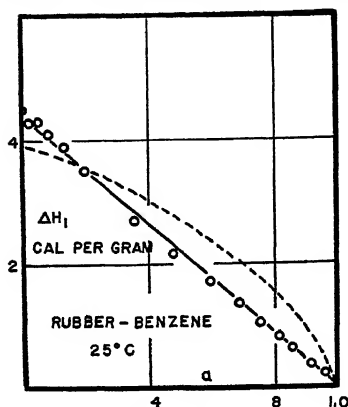


FIGURE 10.

results being shown in FIGURE 10 by the dotted line (as estimated by us from the $\Delta H_0/v_r^2$ function plotted by Gee and Orr).

Inasmuch as the partial molal free energy, $\overline{\Delta F}_1$, of the benzene is given by the expression

$$\overline{\Delta F}_1 = RT \ln a \quad (13)$$

we can derive an equation for $\overline{\Delta F}_1$ valid up to a equal to 0.8 by solving EQUATION 7 for a , i.e.,

$$a = \frac{N}{k(N + N_s)} \quad (14)$$

thus

$$\overline{\Delta F}_1 = RT \ln \frac{N}{N + N_s} - RT \ln k. \quad (15)$$

If we accept the original data of Gee and Treloar as being more correct than those of Gee and Orr (though there is no justification for this procedure), we can write the following equation for the partial molal heat of mixing,

$$\overline{\Delta H}_1 = k' T (1 - a) \quad (16)$$

and by combining EQUATIONS 16 and 15, we obtain an equation for the entropy

$$\overline{\Delta S}_1 = R \ln \frac{N + N_s}{N} + R \ln k + k'(1 - a). \quad (17)$$

EQUATION 17 tells us that, if $R \ln (N + N_s)/N$ with N_s equal to 0.24 is subtracted from Gee and Treloar's tabulated values of $\overline{\Delta S}_1$, the resulting residual entropy should be a linear function of a up to a equal 0.8. FIGURE 11 illustrates this relationship, circles and solid line. The new data of Gee

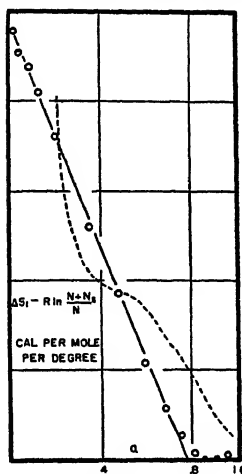


FIGURE 11.

and Orr are shown by the dotted line. Until the uncertainty concerning the partial molal heat of mixing is overcome, we cannot be sure of the correct function for the partial molal entropy of mixing benzene and rubber.

The constant negative entropy term, $R \ln k$, which has the value 0.44, might possibly be explained in terms of the decrease in entropy due to swelling as recently discussed by White and Eyring,¹⁰ provided, of course, that the entropy of swelling per mole of added benzene is constant (White and Eyring's equation gives the free energy of swelling as a function of the volume fraction). We have been unable, as yet, to interpret the constant term $R \ln k$ quantitatively from this point of view.

It is interesting to compare EQUATION 17 with the equation for the entropy of mixing benzene and rubber as derived by Huggins,¹⁶ Flory,¹⁷ and others which is (in its simplest form)

$$\overline{\Delta S_1} = R \left[\ln \frac{N + zN_0}{N} + \frac{zN_0}{zN_0 + N} \cdot \left(1 - \frac{1}{z} \right) \right] \quad (18)$$

where N_0 is the moles of rubber molecules mixed with N moles of benzene and z is the number of segments per single rubber molecule. If we identify zN_0 with N_s , which would be true only approximately, there are certain similarities between EQUATIONS 18 and 17. Or, if our equation for the activity of the benzene

$$\ln a = \ln \frac{N}{N + N_s} - \ln k \quad (19)$$

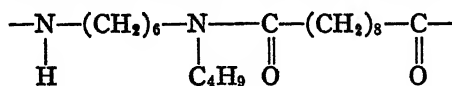
is compared with that of Huggins¹⁸

$$\ln a = \ln v_1 + (1 - \bar{V}_1/\bar{V}_2)v_2 + \mu_1 v_2^2, \quad (20)$$

where v_1 and v_2 are the volume fractions of the benzene and rubber, respectively, V_1 and V_2 are the molal volumes of the pure components and μ is a heat of mixing constant, we see again a certain similarity. At low concentrations of benzene, v_2 becomes nearly equal to unity, and EQUATIONS 20 and 19 become more nearly alike; however, in the case of toluene, where EQUATION 19 is valid up to an activity of toluene equal to 0.95, v_2 approximately 0.5, it is hard to see how the statistical theories of Flory and Huggins can be reconciled with the treatment of this paper. We are open-minded at the present time, however, as to the relative significance of the two points of view.

Sorption of Water by Isobutylated Polyamides

Thanks to the courtesy of the E. I. DuPont de Nemours and Company in supplying us with samples of substituted and unsubstituted polyhexamethylene sebacamides, we have been able to study the sorption of water as a function of *N*-isobutylation. While most of the results will be reserved for a later publication,* we wish to discuss here the sorption of water by the polymer which has, on the average, the following repeating unit:



Data for sorption at 40° are plotted in FIGURE 12 according to the a/N func-

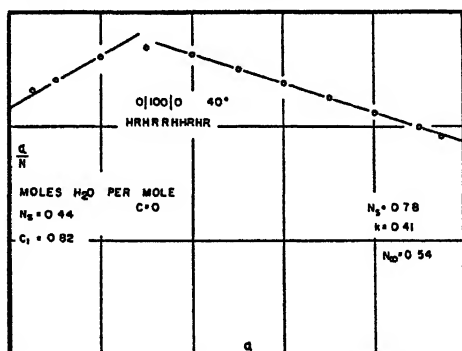


FIGURE 12.

tion; it can be seen at once that the long linear portion from the activity of the water equal to 0.3 up to a equal to 0.95 is more characteristic of the sorption of organic vapor by rubber than of the sorption of water by proteins, nylon, etc. This polymer has a rather low melting point of about 75° C. When water is sorbed, the material becomes somewhat fluid, although quite viscous. Apparently, the polymer partially melts or dissolves in the

* Data obtained by Miss I. L. Fallier.

sorbed water. If the data are plotted according to the BET function (FIGURE 13), good agreement results, particularly if the limited layer BET

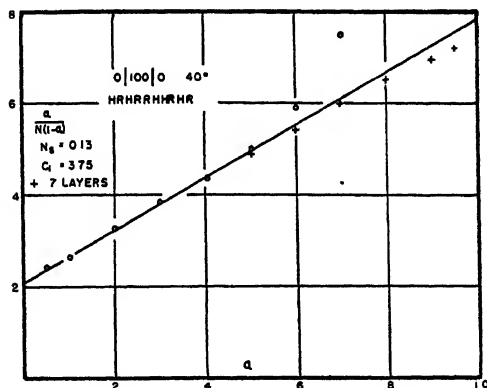


FIGURE 13.

equation is used with N equal to seven. However, seven seems to be too many layers, as judged by the limiting value of a/N when a goes to unity (FIGURE 12). If the BET value of N_s is assumed (0.13), and if this is multiplied by seven, one would expect about 0.91 moles of water sorbed per mole of carbonyl group at saturation. Actually, the data seem to extrapolate to 0.54 moles at saturation.

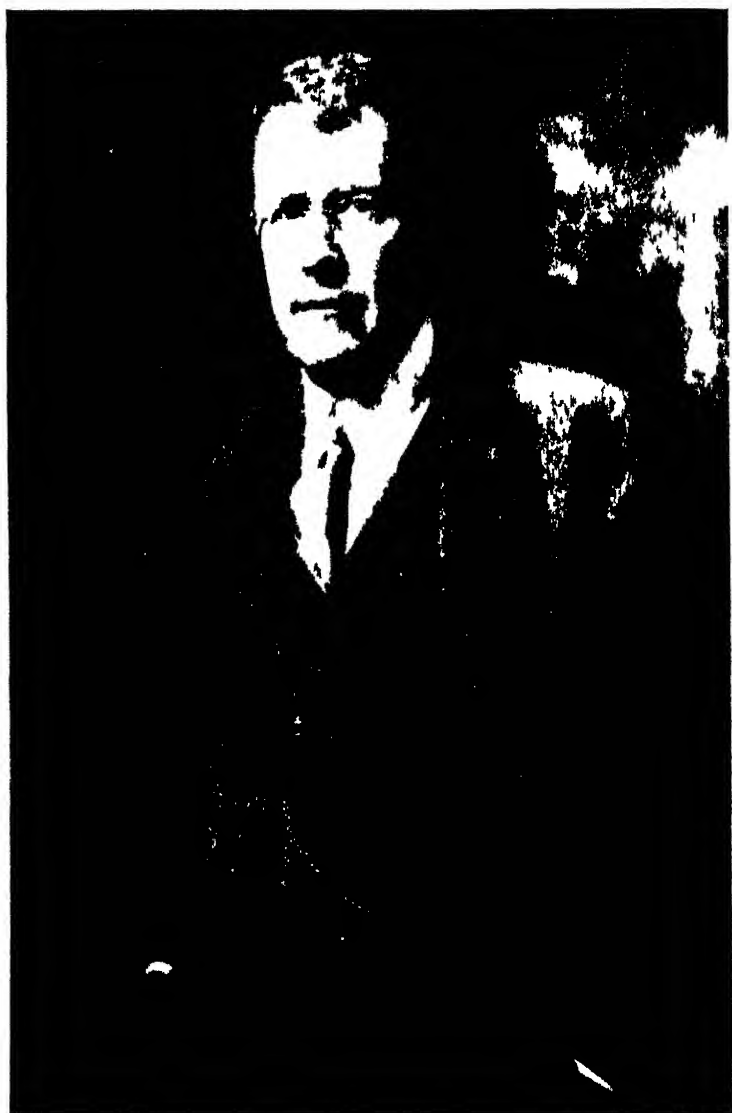
There is very poor agreement between the values of N_s as calculated from the BET equation and from the limiting slopes of FIGURE 12. The Langmuir sorption curve gives N_s equal to 0.44 (almost one water molecule per unsubstituted imino group), while the modified Raoult's law curve suggests 0.78 for N_s (almost one water molecule per carbonyl group). This is a rather meaningless result, as there are apparently only 0.54 moles of water per carbonyl even at saturation.

Experiments on the sorption of water at different temperatures from which the heats of sorption can be calculated are now under way. When they have been completed, the water sorption data for the isobutylated 6-10 nylons will be published in full.

Bibliography

1. FOWLER, R. H., & E. A. GUGGENHEIM. 1939. *Statistical Thermodynamics*: 426. University Press. Cambridge.
2. CASSIE, A. B. D. 1945. *Trans. Faraday Soc.* **41**: 450.
3. HILL, T. L. 1946. *J. Chem. Phys.* **14**: 263.
4. DOLE, M. 1948. *J. Chem. Phys.* **16**: 25.
5. BRUNAUER, S., P. H. EMMETT, & E. TELLER. 1938. *J. Am. Chem. Soc.* **60**: 309.
6. DOLE, M., & A. D. McLAREN. 1947. *J. Am. Chem. Soc.* **69**: 651.
7. BULL, H. B. 1944. Private Communication. See also *J. Amer. Chem. Soc.* **66**: 1499.

8. BARRER, R. M. 1947. Trans. Farad. Soc. **43**: 3.
9. MILLER, A. R. 1942. Proc. Camb. Phil. Soc. **38**: 109.
10. WHITE, H. J., JR., & H. EYRING. 1947. Textile Res. J. **17**: 523.
11. HAUSER, P. M., & A. D. McLAREN. 1948. Ind. Eng. Chem. **40**: 112.
12. LENS, J. 1932. Rec. Trav. Chim. Pays-Bas **51**: 971.
13. MEYER, K. H., E. WOLFF, & CH. G. BOISSONNAS. 1940. Helv. Chim. Acta. **23**: 430.
14. GEE, G., & L. R. G. TRELOAR. 1942. Trans. Farad. Soc. **38**: 147.
15. GEE G. & W. J. C. ORR, Trans. Farad. Soc. **38**: 147, 1942, also follow the $a/N \sim a$ relationship.
16. HUGGINS, M. L. 1942. Ann. N. Y. Acad. Sci. **43**: 1.
17. FLORY, P. J. 1942. J. Chem. Phys. **10**: 51.
18. HUGGINS, M. L. 1943. Ind. Eng. Chem. **35**: 216.



GRINNELL JONES 1884-1947

THE SCIENTIFIC CONTRIBUTIONS OF GRINNELL JONES*

BY MALCOLM DOLE

Department of Chemistry, Northwestern University, Evanston, Illinois

Introduction

It is a very great privilege and pleasure to describe the scientific contributions of Grinnell Jones, Professor of Chemistry at Harvard University from 1912 until his sudden death on June 23, 1947. Above all, Jones had a deep and abiding interest in his students. He gave them every opportunity to discuss with him all their problems, both scientific and personal, and he became essentially a father to them all. The warmth of his personality melted away any "ice curtain" that might have existed between him and a timid student. He was always tremendously interested in and enthusiastic for any new idea which one of his students may have had, and he encouraged his students to think along original lines.

The highest scientific ideals, which were his by inheritance and environment, governed Jones's scientific work and thought and were imparted by him to his many devoted students. Before reviewing his scientific contributions, let us consider the influences which tended to mold his scientific thinking and work.

Influences in Jones's Life

The scholarly and academic tradition was strong in Jones's family. His paternal grandfather, John A. Jones, who married Ann Davis, a school teacher's daughter, became in 1852 a Congregational minister in Berlin, Wisconsin. Jones's father, Richard D. Jones, was a Professor of English Literature, successively, at Syracuse and Vanderbilt Universities and Tufts College. His maternal grandfather, J. B. Grinnell, founded Grinnell College in Iowa. As a boy of seven, Jones was taken to Germany while his father was on a sabbatical leave, and there he acquired a fundamental knowledge of and feeling for the German language which was to be of great help to him in his later scientific career.

Jones graduated *summa cum laude* from Vanderbilt University at the youthful age of nineteen and, after a year as a graduate assistant in both mathematics and chemistry, entered the Harvard Graduate School in the fall of 1904. Jones liked mathematics as well as chemistry, and he once confided to me that had his mathematics teachers at Vanderbilt been able to inspire in him a feeling of the future importance of mathematics, he might have been a mathematician. Jones's ability to handle difficult mathematical treatments stood him in good stead in his unravelling of the mysteries of the Debye-Hückel theory of strong electrolytes when this was first published in 1923.

* The publication of the portrait of Grinnell Jones has been made possible through the generosity of a number of his former students

There is no doubt that Harvard's great physical chemist, T. W. Richards, was the most important single influence on Jones's scientific work. In 1904, when Jones arrived in Cambridge, a young, impressionable youth of twenty, Richards was at the height of his distinguished career. Not only had he at that time become the world's leading authority on atomic weights, he had also begun the study of atomic and molecular volumes and other associated physical properties. Jones threw himself with enthusiastic vigor into Richards's research program, determining the atomic weight of sulfur and measuring the compressibility of a number of halides of sodium, potassium, silver, and thallium. From Richards, Jones not only acquired the detailed attention to the elimination of experimental errors for which Richards was famous, he also absorbed much of his great scientific idealism. The striking similarity between the experimental thinking of Jones and Richards can be seen by reading, for example, the paper of Jones, Taylor, and Vogel published in March, 1948, and comparing its style and treatment of details with the papers on surface tension measurement written by Richards and Coombs, and by Richards and Carver, years before.

It is interesting to note that, because of the interrupting effect of the First World War and of his important work as the Chief Chemist of the U. S. Tariff Commission, Jones's purely scientific contributions were not made in any volume until after 1924, when he was forty years old. In fact, his scientific life "began at forty." That Jones was able to approach the problems he then encountered with an almost boyish enthusiasm can be traced, in part, to his interest in students and in athletics. As a college student, he was captain of the Vanderbilt University track team, and for many years held the Southern Intercollegiate record of $50\frac{2}{3}$ seconds for the 440-yard dash. He never lost this interest in sports, which had the effect, I believe, of helping to maintain his youthful spirit right up to the last days of his life. Furthermore, because of his interest in his students and because of his receptiveness to their suggestions, Jones's thoughts were continually being refreshed by the ideas of young people. He had a positive personality; that is to say, he never discouraged his students from thinking with originality but, on the contrary, he supported and promoted any good ideas which came to him, thus greatly stimulating his students and encouraging further suggestions on their part.

Contributions of Jones to Experimental Techniques

Jones's first great paper was entitled, *The Measurement of the Conductance of Electrolytes I. An Experimental and Theoretical Study of the Principles of Design of the Wheatstone Bridge for Use with Alternating Currents and an Improved Form of Direct-Reading Alternating Current Bridge*. Occupying over forty-nine pages of the Journal of the American Chemical Society, at a time when publication space was severely limited, this paper marked a turning point in the science of conductance measurements. With his able

student, Roswell Colt Josephs, Jones applied new electronic amplifiers and oscillators to this work. He and Josephs discovered errors inherent in the design of resistance boxes, in the method of grounding the bridge, in the use of helical slide wires, in unused but appended resistance coils, and in methods of shielding. They designed a new equal ratio arm bridge, sensitive to one part in a million, new resistance boxes, and a new method of grounding which increased the accuracy of the Wheatstone bridge to a point where it can now be said that uncertainties in the atomic weights of elements are such that solution concentrations cannot be known with any more accuracy than their conductance can be measured. In other words, there is little incentive for further bridge refinements. The "Jones Bridge" is now manufactured and sold by the Leeds and Northrup Company.

Following this initial distinguished paper, a whole series on conductance techniques, numbering nine in all, appeared during the years 1928-1940. Paper II concerned refinements in the oscillator and detector, while Paper III was on the design of cells. Despite the accuracy with which resistance measurements could be made, the cell constant of the conductance cells, which theoretically is independent of the electrolyte used in the cell, appeared to vary as much as 0.33 per cent in certain cases, as Parker¹ had discovered. This effect, known as the "Parker Effect," had baffled electrochemists for a number of years. Jones discovered that the error was due to a capacitive shunt between the electrical leads to the cell and the filling tube, and suggested methods of designing cells so as to overcome this error. His mathematical analysis of the effect left no doubt that adsorption on the electrodes was not the cause, as had been previously suggested.

Paper IV contained a study of the validity of Ohm's Law, while Paper V, on a new method of determining the absolute conductance of potassium chloride solutions, was another long and important paper. Here, the successful idea was to fill a cell with mercury, the material used as the basis for the international definition of the ohm, measure the resistance, then fill the cell with sulfuric acid solutions whose conductance could then be compared to that of the standard potassium chloride solution. In addition to the first paper on this subject, with B. C. Bradshaw, published in 1933, a second paper, with M. J. Prendergast, No. VIII of the series, appeared in 1937. These new conductance values of Jones are now the standard for the world. With S. M. Christian and Dorothy M. Bollinger, Jones studied polarization resistance and polarization capacitance, Paper VI, and the effect of platinization of the electrodes on these properties, Paper VII. His last paper in this series, No. IX, referred to the use of a cathode-ray oscillograph as the detector for the alternating current balance in bridge measurements.

Another important series of papers by Jones and co-workers is to be found in the field of the viscosity of solutions. With S. K. Talley, Jones invented a new photoelectric method for automatically timing the liquid flow in

Ostwald viscometers, which was accurate to 0.01 sec. Previously, attempts to measure the rate of flow were based upon human observation and switch manipulations which, at best, could be reproducible to only about 0.06 to 0.1 sec. By eliminating psychological errors, Jones obtained data on the viscosity of a number of electrolytes which are unsurpassed. He also investigated the possible influence of surface tension and surface drainage upon viscosity measurements.

With W. A. Ray, Jones constructed a new differential tensiometer for the accurate measurement of the relative surface tension of strong electrolytes in order to test the validity of the Debye-Hückel theory in this realm. His idea was to bring the surface of the liquid in the capillary tube always to the same point and to measure the differences in the height of rise by weighing the amount of solution required in the side vessel to bring the liquid to the desired point in the capillary tube. By using carefully constructed all-quartz apparatus, Jones and Ray were able to measure surface tension with the amazing relative sensitivity of 0.001 per cent. This work led to the discovery of the "Jones-Ray Effect" discussed later and to the compilation of many accurate surface tension values.

Other inventions of Jones in the field of experimental techniques include an "equilibrator," a device for bringing two separate liquids to equilibrium with respect to the same vapor phase, invented in 1928; a new dilatometer (1935) which employed the unique principle of being its own thermal regulator; and an improved apparatus for the measurement of interfacial potentials at the interface between vitreous silica and aqueous solutions (1945).

Theoretical Contributions of Jones

The first paper to be published by Jones after obtaining his Ph.D. was entitled *An Explanation of the Negative Coefficient of Expansion of Silver Iodide*, in which he showed that a free energy of formation of silver iodide which increased with temperature was fundamentally responsible for the unique property of silver iodide to contract on heating.

In 1929, Jones published a paper on the transference number of barium in barium chloride solutions, in which he applied a mathematical equation to the description of the results, this equation being one of the first in the field of transference numbers. Also in 1929 came perhaps his most important theoretical work, the application of the ideas of Debye and Hückel to the new field of viscosity of strong electrolytes. Jones showed that a square root term existed in the viscosity equation for salt solutions, a term for which Falkenhagen and Dole² and Falkenhagen and Vernon³ later supplied the mathematical theory. Furthermore, Jones and Dole were able to predict that the viscosity of potassium chloride solutions, for example, which is less than that of water at moderate concentrations, must become greater than the viscosity of water at very low concentrations. This maximum in the viscosity-concentration curve had never been observed, but the

English workers, Joy and Wolfenden,⁴ and others, soon verified the prediction. The viscosity data obtained by Jones and his co-workers have verified brilliantly the major postulates of the modern theory of strong electrolytes.

In 1935, Jones and Ray unexpectedly discovered a minimum in the surface tension-concentration curve of all strong electrolytes, as measured by their new quartz tensiometer. Amounting only to a 0.02 per cent decrease in surface tension, it was thought by Langmuir⁵ to be a zeta-potential effect. Langmuir suggested that the initial decrease in surface tension, which violated the predictions of the Debye-Hückel theory, resulted from a disappearance on adding electrolyte of a surface film held on the quartz by zeta-potential forces. To test this explanation of Langmuir's, accurate zeta-potential values were required, a need which led Jones and Wood to the development of the refined zeta-potential apparatus mentioned previously. Although some of the consequences of Langmuir's theory seem to be fulfilled, further work in this field may lead to new light on the important subject of interfacial potentials.

On looking over the many highly accurate and unsurpassed data in the field of electrical conductances, viscosity, surface tension, and expansibilities of solutions of electrolytes, one realizes that Jones's results will be a mine in which the theoreticians may dig with great profit and understanding for many years to come.

In conclusion, it should be pointed out that Grinnell Jones's inventiveness was not limited entirely to strictly scientific research problems. In the field of practical inventions, he devised, first, a new type of silver polish; second, a method of impregnating cloth with silver nitrate so that the cloth would protect silverware wrapped within it from tarnishing; and, third, a new type of fire-resistant paint. Used exclusively by the Army during the last World War, this new anti-fire paint can now be applied by the general public for the protection against fire of all wooden structures. The future saving to the country may be immense.

References Cited

1. PARLER, H. C. 1923. *J. Am. Chem. Soc.* **45**: 1366, 2017.
2. FALKENHAGEN, H., & M. DOLE. 1929. *Phys. Z.* **30**: 611.
3. FALKENHAGEN, H., & E. L. VERNON. 1932. *Phil. Mag.* [7] **14**: 537.
4. JOY, W. E., & J. H. WOLFENDEN. 1931. *Proc. Roy. Soc. London A* **134**: 413.
5. LANGMUIR, I. 1938. *Science* **88**: 430. 1938, *J. Chem. Phys.* **6**: 894.

Bibliography

(References to Paper on the Tariff and to Patents Omitted)

1907. The molecular weight of silver sulphate and the atomic weight of sulfur. (With T. W. RICHARDS.) *J. Am. Chem. Soc.* **29**: 826; *Z. anorg. Chem.* **55**: 72.
1909. The compressibilities of the chlorides, bromides, and iodides of sodium, potassium, silver and thallium. (With T. W. RICHARDS.) *J. Am. Chem. Soc.* **31**: 158.
- 1909-1910. An explanation of the negative coefficient of expansion of silver iodide. *J. Am. Chem. Soc.* **31**: 191; *Z. phys. Chem.* **71**: 179.

1910. A revision of the atomic weight of phosphorus I. The Analysis of silver phosphate. (With G. P. BAXTER.) *Proc. Am. Acad. Arts & Sci.* **45**: 138; *J. Am. Chem. Soc.* **32**: 298.
1910. The atomic weight of hydrogen. *J. Am. Chem. Soc.* **32**: 513.
1915. Study of the system: water, potassium iodide and iodine at zero degrees. (With M. L. HARTMANN.) *J. Am. Chem. Soc.* **37**: 241.
1915. Properties of silver iodide interpreted in relation to recent thermodynamic conceptions. (With M. L. HARTMANN.) *J. Am. Chem. Soc.* **37**: 752.
1916. The equilibrium between bromine and potassium bromide solutions at zero degrees. (With M. L. HARTMANN.) *Trans. Am. Electrochem. Soc.* **30**: 295.
1921. Potential of the thallium electrode and the free energy of formation of thallous iodide. (With W. C. SCHUMB.) *Proc. Am. Acad. Arts & Sci.* **56**: 199.
1928. The measurement of the conductance of electrolytes I. An experimental and theoretical study of principles of design of the Wheatstone bridge for use with alternating currents and an improved form of direct-reading alternating current bridge. (With R. C. JOSEPHS.) *J. Am. Chem. Soc.* **50**: 1049.
1928. A new equilibrator: a device for the determination of the distribution ratio of a volatile solute between two miscible solvents. (With B. B. KAPLAN.) *J. Am. Chem. Soc.* **50**: 1600.
1928. Normal potential of the iodine-Iodide electrode. (With B. B. KAPLAN.) *J. Am. Chem. Soc.* **50**: 2066.
1928. The iodide, iodine, tri-iodide equilibrium and the free energy of formation of silver iodide." (With B. B. KAPLAN.) *J. Am. Chem. Soc.* **50**: 1845.
1929. The transference number of barium chloride as a function of the concentration. (With MALCOLM DOLE.) *J. Am. Chem. Soc.* **51**: 1073.
1929. The measurement of the conductance of electrolytes II. Improvements in the oscillator and detector. (With GILES M. BOLLINGER.) *J. Am. Chem. Soc.* **51**: 2407.
1929. Viscosity of aqueous solutions of strong electrolytes with special reference to barium chloride. (With MALCOLM DOLE.) *J. Am. Chem. Soc.* **51**: 2950.
1930. On the existence and behavior of complex polyiodides. *J. Phys. Chem.* **34**: 673.
1930. The electrical conductance of aqueous solutions of barium chloride as a function of the concentration. (With MALCOLM DOLE.) *J. Am. Chem. Soc.* **52**: 2245.
1931. Formation of films at liquid-liquid interfaces. (With J. A. SERRALLACH.) *Ind. Eng. Chem.* **23**: 1016.
1931. The measurement of the conductance of electrolytes III. The design of cells. (With GILES M. BOLLINGER.) *J. Am. Chem. Soc.* **52**: 411.
1931. The measurement of the conductance of electrolytes IV. The validity of Ohm's Law for electrolytes. (With GILES M. BOLLINGER.) *J. Am. Chem. Soc.* **53**: 1207.
1932. The transference number of lithium chloride as a function of the concentration. (With BENJAMIN C. BRADSHAW.) *J. Am. Chem. Soc.* **54**: 138.
1933. Strength of emulsifier films at liquid-liquid interfaces. (With J. A. SERRALLACH & R. J. OWEN.) *Ind. Eng. Chem.* **25**: 816.
1933. The viscosity of aqueous solutions as a function of the concentration. (With SAMUEL K. TALLEY.) *J. Am. Chem. Soc.* **55**: 624.
1933. Automatic timing of the Ostwald viscometer by means of a photoelectric cell. (With S. K. TALLEY.) *Physic.* **4**: 215.
1933. The measurement of the conductance of electrolytes V. A redetermination of the conductance of standard potassium chloride solutions in absolute units. (With BENJAMIN C. BRADSHAW.) *J. Am. Chem. Soc.* **55**: 1780.
1933. The viscosity of aqueous solutions as a function of concentration II. Potassium bromide and potassium chloride. (With SAMUEL K. TALLEY.) *J. Am. Chem. Soc.* **55**: 4124.
1934. The conductance of aqueous solutions as a function of concentration I. Potassium bromide and lanthanum chloride. (With CHARLES F. BICKFORD.) *J. Am. Chem. Soc.* **56**: 602.
1934. Equilibria in aqueous solutions of bromine and potassium bromide. (With SVEN BAECKSTROM.) *J. Am. Chem. Soc.* **56**: 1517.
1934. The standard potential of the bromine electrode. (With SVEN BAECKSTROM.) *J. Am. Chem. Soc.* **56**: 1524.
1935. The measurement and conductance of electrolytes VI. Galvanic polarization by alternating current. (With SCHUYLER M. CHRISTIAN.) *J. Am. Chem. Soc.* **57**: 272.
1935. The measurement of the conductance of electrolytes VII. Platinization. (With DOROTHY M. BOLLINGER.) *J. Am. Chem. Soc.* **57**: 280.

1935. Surface tension of solutions. (With WENDELL A. RAY.) *J. Am. Chem. Soc.* **57**: 957.
1935. Viscosity of solutions of salts in methanol. (With HOLMES J. FORNWALT.) *J. Am. Chem. Soc.* **57**: 2041.
1935. The coefficient of expansion of silver iodide and of the halides of thallium. (With FREDERIC C. JELEN.) *J. Am. Chem. Soc.* **57**: 2532.
1936. The viscosity of aqueous solutions as a function of the concentration III. Cesium iodide and potassium permanganate. (With HOLMES J. FORNWALT.) *J. Am. Chem. Soc.* **58**: 619.
1936. The transference number of lanthanum chloride as a function of concentration. (With LAURENCE T. PRENDERGAST.) *J. Am. Chem. Soc.* **58**: 1476.
1936. The viscosity of deuterium oxide and its mixtures with water at 25°. (With HOLMES J. FORNWALT.) *J. Chem. Physics* **4**: 30.
1936. The conductance of aqueous solutions as a function of the concentration II. potassium ferrocyanide (With FREDERIC C. JELEN.) *J. Am. Chem. Soc.* **58**: 2561.
1936. Viscosity of aqueous solutions as a function of the concentration IV. Potassium ferrocyanide. (With ROBERT E. STAUFFER.) *J. Am. Chem. Soc.* **58**: 2558.
1937. Surface tension of solutions of electrolytes as a function of the concentration I. A differential method for measuring relative surface tensions. (With WENDELL A. RAY.) *J. Am. Chem. Soc.* **59**: 187.
1937. The measurement of the conductance of electrolytes VIII. A redetermination of the conductance of Kohlrausch's standard potassium chloride solutions in absolute units. (With MAURICE J. PRENDERGAST.) *J. Am. Chem. Soc.* **59**: 731.
1937. The surface tension of deuterium oxide and its mixtures with water. (With WENDELL A. RAY.) *J. Chem. Physics* **5**: 505.
1937. The viscosities of aqueous solutions of electrolytes as functions of the concentration V. Sodium chloride. (With SCHUYLER M. CHRISTIAN.) *J. Am. Chem. Soc.* **59**: 484.
1937. Drainage error in viscometry of aqueous solutions. (With ROBERT E. STAUFFER.) *J. Am. Chem. Soc.* **59**: 1630.
1938. The influence of surface tension on the measurement of viscosity. The viscosity of methanol. (With HOLMES J. FORNWALT.) *J. Am. Chem. Soc.* **60**: 1683.
1939. The viscosity of normal butanol. (With S. M. CHRISTIAN.) *J. Am. Chem. Soc.* **61**: 82.
1939. The drainage of viscometers and pipettes. (With EDNA FERRELL.) *J. Am. Chem. Soc.* **61**: 325.
1940. The viscosity of solutions of electrolytes as a function of the concentration VI. Potassium bromide and lanthanum chloride. (With ROBERT E. STAUFFER.) *J. Am. Chem. Soc.* **62**: 335.
1940. The viscosity of solutions of electrolytes as a function of the concentration VII. Silver nitrate, potassium sulfate and potassium chromate. (With JOHN H. COLVIN.) *J. Am. Chem. Soc.* **62**: 338.
1940. The measurement of the conductance of electrolytes IX. The use of a cathode-ray oscillograph as a detector. (With KAROL J. MYSELS & WALTER JUDA.) *J. Am. Chem. Soc.* **62**: 2919.
1940. A theoretical and experimental analysis of the capillary rise method for measuring the surface tension of solutions of electrolytes. (With LAURENCE D. FRIZZELL.) *J. Chem. Physics* **8**: 986.
1941. The surface tension of solutions of electrolytes as a function of the concentration II. (With WENDELL A. RAY.) *J. Am. Chem. Soc.* **63**: 288.
1941. The surface tension of solution of electrolytes as a function of the concentration III. Sodium chloride. (With WENDELL A. RAY.) *J. Am. Chem. Soc.* **63**: 3262.
1942. The surface tension of solutions of electrolytes as a function of the concentration IV. Magnesium sulfate. (With WENDELL A. RAY.) *J. Am. Chem. Soc.* **64**: 2744.
1944. The viscosity of aqueous solutions of electrolytes as a function of the concentration VIII. Potassium Ferricyanide. (With SCHUYLER M. CHRISTIAN.) *J. Am. Chem. Soc.* **66**: 1017.
1944. Electrochemical studies on vanadium salts I. The vanadyl-vanadic oxidation-reduction potential. (With JOHN H. COLVIN.) *J. Am. Chem. Soc.* **66**: 1563.
1944. Electrochemical studies on vanadium salts II. The hydrolysis of vanadyl sulfate and vanadic sulfate solutions. (With WENDELL A. RAY.) *J. Am. Chem. Soc.* **66**: 1571.
1944. Electrochemical studies on vanadium salts III. The vanadic-vanadous oxidation-reduction potential. (With JOHN H. COLVIN.) *J. Am. Chem. Soc.* **66**: 1573.

1945. The measurement of potentials at the interface between vitreous silica and solutions of potassium chloride by the streaming potential method. (With LLOYD A. WOOD.) *J. Chem. Physics* 13: 106.
1948. The apparent molal expansibilities of potassium, barium and lanthanum chlorides as a function of concentration and temperature. (With E. FOLGER TAYLOR & RICHARD C. VOGEL.) *J. Am. Chem. Soc.* 70: 966.

Former Students and Research Associates of Jones

- SVEN BAECKSTROM, Ph.D. Chemist, Hantverkaregatan 3, Hofors, Sweden.
- C. F. BICKFORD, Ph.D. Chemist, E. R. Squibb and Sons.
- DOROTHY M. BOLLINGER. Women's Medical College, Philadelphia, Pa.
- G. M. BOLLINGER, Ph.D. Deceased (Formerly Assistant Professor of Chemistry at Bowdoin College, Maine).
- B. C. BRADSHAW, Ph.D. Pembroke, Ky.
- G. E. CHALLENGER, Ph.D. Los Alamos, N. Mex.
- S. M. CHRISTIAN, Ph.D. Agnes Scott College, Decatur, Ga.
- HARVEY SHIELDS COLLINS, A.M., M.D. Research Fellow in Medicine, Harvard University Medical School, and Thorndike Memorial Hospital, Boston, Mass.
- J. H. COLVIN, Ph.D. Patent Department, Shell Development Co., San Francisco, Calif.
- THOMAS WENDELL DAKIN, Ph.D. Group Leader, Westinghouse Research Laboratory, E. Pittsburgh, Penna.
- M. DOLE, Ph.D. Professor of Chemistry, Northwestern University, Evanston, Illinois.
- EDNA FERRELL (Mrs. Caryl P. Haskins). Research Staff, The Haskins Laboratories, New York, N. Y.
- H. J. FORNWALT, Ph.D. Chemical Engineer, Sales Division, Darco Corporation, New York, N. Y.
- L. D. FRIZZELL, Ph.D. Assistant Professor of Chemistry, Northwestern University, Evanston, Ill.
- M. L. HARTMANN, Ph.D. Consulting Chemical Engineer, Beverly Hills, Calif.
- F. C. JELEN, Ph.D. Chemical Engineer, Cowles Detergent Co., Cleveland, Ohio.
- R. C. JOSEPHS, 98 Lexington St., Weston, Mass.
- WALTER JUDA, A.M. Instructor in Chemistry, Harvard University, Cambridge, Mass.
- B. B. KAPLAN, A.M. President, Double B Products Co.; Vice-President, Albi Manufacturing Co., Hartford, Conn.
- ANDREW A. KASPER, A.M. Instructor in Chemistry, Boston University, Boston, Mass.
- WAYNE ALAN MCRAE. Graduate Student in Chemistry, Harvard University, Cambridge, Mass.
- KAROL J. MYSELS, Ph.D. Assistant Professor of Chemistry, University of Southern California, Los Angeles, Calif.
- RANDOLPH J. OWEN. Production Manager, The E. L. Patch Company, Stoneham, Mass.
- L. T. PRENDERGAST, Ph.D. Molded Latex Products, Inc. Paterson, N. J.
- M. J. PRENDERGAST, Ph.D. E. H. Sargent Co., Chicago, Ill.
- W. A. RAY, Ph.D. Professor of Chemistry, Colby College, Waterville, Maine.
- W. C. SCHUMB, Ph.D. Professor of Inorganic Chemistry, Massachusetts Institute of Technology, Cambridge, Mass.
- J. A. SERRALLACH. Chemical Engineer, Barcelona, Spain.
- R. E. STAUFFER, Ph.D. Research Laboratories, Eastman Kodak Co., Rochester, N. Y.
- E. F. TAYLOR, A.M. Research Engineer, E. I. DuPont de Nemours & Co., Waynesboro, Va.
- R. C. VOGEL, Ph.D. Assistant Professor of Chemistry, Illinois Institute of Technology, Chicago, Ill.
- L. A. WOOD, Ph.D. Assistant Professor of Chemistry, Cornell University, Ithaca, N. Y.

PRESSURE-VOLUME-TEMPERATURE RELATIONS IN SOLUTIONS VIII. THE BEHAVIOR OF SOME SOLUTIONS OF ELECTROLYTES IN WATER

By R. E. GIBSON* AND O. H. LOEFFLER†

Geophysical Laboratory, Carnegie Institution of Washington, Washington, D. C.

Introduction. A completely satisfactory theory of molecular interaction as applied to solutions should enable us to predict with certainty at any temperature, pressure, and composition the thermodynamic properties of a solution from a knowledge of definable macroscopic or microscopic properties of the pure components. This is the long range objective of the study of molecular interaction in solutions and requires a knowledge not only of the molecules and the forces between them but also of the distribution functions and other quantities needed in any statistical treatment. Up to the present time, the theory of liquids and liquid solutions has been in such an undeveloped state that fruitful attacks could be made only on more limited objectives; and empirical work on the effect of one of the variables, pressure, temperature, or concentration on specific properties such as volume, chemical potential, *etc.*, has been in order. Even now, in the case of the most thoroughly investigated systems, namely, solutions of electrolytes in water, the theory extends little beyond an exact formulation of the limiting laws and second approximations thereto, and the experimental data available do not cover a wide range of conditions.

Nearly ten years ago, we undertook a systematic study of the volume changes in electrolytic solutions over a considerable range of concentration, pressure, and temperature in order to accumulate data for a broader consideration of the effects of molecular interaction on volume changes in these systems. This work was part of a study of the effect of pressure and temperature on the solubility of inorganic substances in water and, therefore, emphasized the more concentrated solutions. Little attention was paid to very dilute solutions, which are adequately treated in other papers presented at this conference. Solutions of *sodium bromide*, *sodium chloride*, and *lithium bromide*, covering the whole range of concentration, were examined between 25° and 85° and over the pressure range of 1 to 1,000 bars. The results are complicated and a condensed presentation of them does not introduce simplification. However, an attempt will be made in this paper to summarize in a usable form the results for sodium bromide and sodium chloride in water and to indicate some of the more obvious significant conclusions that may be drawn from the data. In order to clarify the results obtained for the water solutions, reference will be made, first, to similar measurements on a simpler binary system, namely sodium bromide dissolved in ethylene glycol,

* Present address: Applied Physics Laboratory, The Johns Hopkins University, Silver Spring, Maryland.

† Present address: Bureau of Ordnance, U. S. Navy, Washington, D. C.

and to the behavior of pure liquids under changes of pressure and temperature.

Experimental and Derived Quantities. The primary experimental data consist of accurate measurements of the specific volumes of the solutions at 25° and 1 atmosphere, the expansions from 25° to a series of temperatures separated by 10° intervals up to 85°, and the compressions from 1 bar to 500 and 1000 bars at the same temperatures. The techniques employed for these measurements of specific volumes,¹ expansions,² and compressions³ have already been described.

Volume-Temperature Relations. Every effort was made to express the thermal expansions as functions of temperature by equations which fitted the data over a wide range of temperature with sufficient accuracy to insure that the temperature coefficients (thermal expansibilities) computed from the equations were as good as the original measurements permitted. In general, a cubic equation of the form

$$v = v_{25} + a(t - 55)^2 + b(t - 55)^2 + c(t - 55)^3 \quad (1)$$

was found to be adequate, t being the temperature in °C, and v the specific volume. However, in certain cases, it was found necessary to have recourse to a more complicated equation of the form originally proposed by Ipatov. We have already discussed the application of these equations to aqueous solutions and presented the results for the expansion coefficients of the solutions of sodium bromide and sodium chloride which are the subject of this paper.⁴

Volume-Pressure Relations—The Tait Equation. The systematization of volume-pressure relations for liquids and solutions has been greatly simplified by the application of the equation proposed by P. G. Tait in 1880. According to this relation,

$$-\frac{\Delta_p v}{v_0} = C \log \left(\frac{B + P}{B + P_0} \right) \quad (2)$$

where $-\Delta_p v$ is the compression when the pressure changes from P_0 to P ; v_0 is the volume at P_0 ; C is a constant which is independent of temperature, and B is a constant which depends on temperature. This equation represents the compressibility results for a wide variety of liquids with a remarkable degree of precision and, indeed, has been used in a modified form up to 25,000 atmospheres.⁵ Furthermore, it has been found that the constant C has exactly the same value for similar liquids such as benzene and its derivatives⁶ or for solutions in a given solvent.⁷ Up to the present it has not been found possible to relate the Tait equation with any theoretical considerations concerning liquids, but its remarkable power of expressing experimental data suggests that study in this direction would be most profitable. By means of the Tait equation, it is possible to obtain the volume and its pressure derivatives as a function of pressure at any temperature. Furthermore, by

differentiating the Tait equation with respect to temperature, it was possible to calculate the expansion coefficients at any pressure from a knowledge of the temperature coefficient of the quantity B .⁸

Thermodynamic Functions. As a result of these measurements and calculations, it is possible to obtain for each liquid or homogeneous solution the equivalent of a table of volumes at different pressures and temperatures and also a table of such derivatives* as

$$\left(\frac{\partial V}{\partial T}\right)_P, \left(\frac{\partial V}{\partial P}\right)_T, \left(\frac{\partial P}{\partial T}\right)_V \equiv -\left(\frac{\partial V}{\partial T}\right)_P / \left(\frac{\partial V}{\partial P}\right)_T, \left(\frac{\partial E}{\partial V}\right)_T \equiv \left(\frac{\partial P}{\partial T}\right)_V - P$$

and $\left(\frac{\partial C_V}{\partial V}\right)_T \equiv \left[\partial \left(\frac{\partial E}{\partial V}\right)_T / \partial T\right]_V$.

Volume Change on Mixing. In considering solutions, the volume change that occurs when the components are mixed is a matter of importance which is usually expressed in terms of the apparent volume of the dissolved substance, ϕ_2 , or the apparent molal volume, Φ_2 . The apparent volume is defined by the expression

$$\phi_2 = \frac{v}{x_2} - \frac{x_1}{x_2} v_1^0 \quad (3)$$

$$\Phi_2 = M_2 \phi_2 \quad (4)$$

where v is the specific volume of the solution, x_1 and x_2 are the weight fractions of solvent and solute respectively, and v_1^0 and v_2^0 the specific volumes of the pure solvent and solute. It will be seen that $(v_2^0 - \phi_2)$ is the contraction that takes place when 1 gram of solute is mixed with sufficient solvent to form a solution of a given concentration. For several series of solutions, ϕ_2 and its temperature and pressure derivatives were computed for any temperature, pressure or concentration in the range studied.

Results. The expansion coefficients of a series of sodium chloride and sodium bromide solutions in water at ordinary pressure have already been published.⁹ The same paper also records equations for expressing the specific and apparent volumes as functions of concentration, thereby permitting the calculation of the various quantities at any concentration. The compressions and compressibilities may be computed from the Tait equation, with the values of B or $B + P_e$ † given in TABLES 2 and 1. It will be noted that

* The symbols are used here with their usual meaning; i.e. V =total volume, E =total energy, P =total pressure, T =temperature, and C_V =heat capacity at constant volume.

† In later work, it became our practice to apply equation 2 directly to solutions using C for the pure solvent and computing B' the value of B for the solution from one or two observations.¹⁰ However, in an earlier work under the influence of the Tamman hypothesis, we treated all solutions as modifications of water and derived a quantity P_e for solution by an equation of the following form:

$$\Delta_p v = C_m v_1^0 \log \frac{B + P_e + P}{B + P_e} - m \Delta_p v^0 \quad (5)$$

Roughly speaking, $(B + P_e)$ corresponds to B' . Agreement with experiment does not indicate a preference between these two applications of the Tait equation.

TABLE 1
VALUES OF $(B + P_e)$ IN EQUATION 5
 $C = 0.3150$

x_2	$(B + P_e)$ in kilobars						
	25°	35°	45°	55°	65°	75°	85°
Sodium Chloride—Water							
0.04946	2.996	3.055	3.081	3.078	3.052	3.005	2.939
.10187	3.267	3.307	3.316	3.305	3.277	3.223	3.155
.14890	3.558	3.578	3.572	3.550	3.515	3.455	3.383
.19957	3.832	3.834	3.811	3.779	3.735	3.669	3.593
.24929	4.139	4.122	4.075	4.030	3.976	3.903	3.821
	4.433	4.400	4.327	4.276	4.205	4.132	4.034
Sodium Bromide—Water							
0.04906	3.123	3.171	3.186	3.180	3.149	3.101	3.035
.14997	3.405	3.426	3.424	3.402	3.360	3.307	3.241
.25004	3.731	3.721	3.695	3.652	3.599	3.538	3.469
.34554	4.092	4.044	3.989	3.923	3.856	3.782	3.703
.44338	4.479	4.396	4.311	4.214	4.136	4.037	3.938

TABLE 2
COEFFICIENTS USED IN DIRECT APPLICATION OF EQUATION 2 TO SOLUTIONS OF SODIUM BROMIDE IN WATER
 $C = 0.3150$

x_2	B in kilobars			
	25°	45°	65°	85°
0.04906	3.158	3.226	3.188	3.080
.14497	3.549	3.577	3.513	3.383
.25004	4.021	3.995	3.898	3.757
.34554	4.577	4.485	4.344	4.181
.44338	5.290	5.122	4.937	4.707

the constant C in the Tait equation is taken as 0.3150, the best value obtained for water. The variation of $(B + P_e)$ with temperature was expressed by equations of the following form.

$$(B + P_e) = (B + P_e)_{55} + \alpha(t - 55) + \beta(t - 55)^2 + \gamma(t - 55)^3 \quad (6)$$

and coefficients for the different solutions are recorded in TABLE 3.

The thermodynamic functions computed from the original data are recorded in TABLES 4 to 13.

Discussion of Results. It has not yet been possible for us to make a complete analysis of the foregoing experimental results in order to interpret the effects of pressure, temperature, and concentration on the various thermodynamic properties in terms of molecular interaction and geometry. In this paper, we shall discuss the variation of certain properties with volume and with temperature, in order to determine the relative effects of these vari-

TABLE 3
 COEFFICIENTS IN EQUATION 6 EXPRESSING $(B + P_2)$ AS FUNCTIONS OF TEMPERATURE

x_2	$(B + P_2)_{65}$	$10^3\alpha$	$10^6\beta$	$10^7\gamma$
<i>NaBr—H₂O</i>				
0.04906	3.178	-1.944	-10.70	7.22
.14497	3.401	-3.205	-8.69	5.28
.25004	3.653	-4.788	-5.83	4.72
.34554	3.925	-6.631	-3.00	1.667
.44338	4.221	-8.850	-1.33	-1.94
<i>NaCl—H₂O</i>				
0.04946	3.307	-2.153	-10.58	3.06
.10187	3.552	-3.054	-9.00	1.39
.14890	3.780	-4.050	-7.49	.56
.19957	4.032	-5.325	-5.63	0
.24929	4.276	-6.450	-4.34	-2.5

ables. It will be assumed that properties which depend only on volume measure effects arising from long range forces in the liquid. By long-range forces are meant those which depend primarily on the mean distance between molecules and which are not greatly influenced by the distribution within this mean distance. Ion-dipole, ion-ion forces, Van der Waal's attractive forces are examples. On the other hand, it is expected that effects arising from short-range or directed forces such as repulsion forces, hydrogen bonds, *etc.*, will be reflected in thermodynamic quantities that vary with temperature when the volume is kept constant. The quantities to which we wish to give special attention are the energy-volume coefficient $(\partial E/\partial V)_T$, the heat capacity at constant volume C_V , the apparent volume of a dissolved substance, ϕ_2 , and the constant B in the Tait equation.

Behavior of Pure Liquids. Although it has often been taken for granted that the energy-volume coefficient $(\partial E/\partial V)_T$ is a pure volume function in normal liquids, we have shown that even in liquids such as carbon tetrachloride¹¹ and benzene¹² it shows a significant temperature dependence, as may be seen in FIGURES 1 and 2. At any given volume, $(\partial E/\partial V)_T$ for these liquids drops as the temperature is raised, indicating that the specific heat at constant volume increases as the liquid is compressed. The increase in repulsive potential energy arising from the inter-penetration of compressible molecules gives an explanation for this effect.¹¹ We have applied the term, "hindered translation," to this phenomenon to signify a potential energy associated with the translational degrees of freedom that arises from the compression of the molecules.¹¹ For carbon tetrachloride,¹¹ benzene, and a number of its derivatives,¹² we found that a quantity obtained by combining $(\partial E/\partial V)_T$ with the constant B in the Tait equation, namely $[(\partial E/\partial V)_T + B + P = T\gamma + B]$, is a pure volume function over a temperature and pressure range that corresponds to nearly 14 per cent change in volume. In

TABLE 4
 SODIUM CHLORIDE—WATER (5 PER CENT)

	$x_2 = 0.04946$		$x_1 = 0.95054$		$X_2 = 0.0158^*$			
	P bars	$\frac{10^4}{v_0} \left(\frac{dv}{dt} \right)_P$ deg ⁻¹	$\frac{10^4}{v_0} \left(\frac{dv}{dP} \right)_T$ bar ⁻¹	$\left(\frac{dP}{dT} \right)_V$ bars/deg.	$\left(\frac{dE}{dV} \right)_T$ bars	v cc.	V^\dagger cc.	ϕ_2 cc.
25°	1	3.203	0.4129	7.757	2311	0.96889	18.07	0.3146
	250	3.323	.3837	8.660	2332	.95928	17.89	.3306
	500	3.429	.3583	9.570	2353	.95031	17.73	.3449
	750	3.517	.3362	10.46	2368	.94190	17.57	.3571
	1000	3.596	.3164	11.37	2389	.93400	17.42	.3678
35°	1	3.840	.4077	9.419	2901	.97231	18.14	
	250	3.876	.3793	10.22	2899	.96279	17.96	
	500	3.906	.3544	11.02	2895	.95388	17.79	
	750	3.930	.3326	11.82	2892	.94553	17.64	
	1000	3.949	.3134	12.60	2882	.93769	17.49	
45°	1	4.415	.4065	10.86	3454	.97634	18.21	0.3324
	250	4.377	.3781	11.58	3434	.96681	18.04	.3461
	500	4.339	.3535	12.27	3403	.95789	17.87	.3577
	750	4.303	.3319	12.96	3373	.94952	17.71	.3676
	1000	4.271	.3127	13.66	3345	.94167	17.57	.3769
55°	1	4.949	.4077	12.14	3982	.98092	18.30	
	250	4.840	.3793	12.76	3937	.97131	18.12	
	500	4.741	.3545	13.37	3887	.96233	17.95	
	750	4.653	.3326	13.99	3840	.95390	17.79	
	1000	4.572	.3134	14.59	3787	.94599	17.65	
65°	1	5.455	.4112	13.27	4486	.98604	18.39	0.3374
	250	5.279	.3823	13.81	4419	.97630	18.21	.3508
	500	5.123	.3571	14.35	4352	.96720	18.04	.3623
	750	4.983	.3350	14.87	4278	.95867	17.88	.3722
	1000	4.856	.3154	15.40	4207	.95066	17.73	.3813
75°	1	5.941	.4182	14.21	4946	.99167	18.50	
	250	5.699	.3883	14.68	4860	.98172	18.31	
	500	5.485	.3623	15.14	4770	.97243	18.14	
	750	5.295	.3395	15.60	4680	.96373	17.98	
	1000	5.124	.3196	16.03	4580	.95557	17.83	
85°	1	6.411	.4273	15.00	5371	.99781	18.61	0.3336
	250	6.102	.3962	15.40	5265	.98759	18.42	.3476
	500	5.831	.3691	15.80	5158	.97805	18.25	.3597
	750	5.590	.3456	16.17	5040	.96914	18.08	.3702
	1000	5.375	.3248	16.55	4927	.96079	17.92	.3793

* X_2 is the mole fraction of salt.† V is the molal volume of the solution.

the above expression, γ denotes $(\partial P / \partial T)_V$ and should not be confused with γ used in EQUATION 6. These results are illustrated in FIGURES 3 and 4. Indeed, it has been shown¹¹ that this quantity $(T\gamma + B)$ may be expressed in the form A/V^3 for carbon tetrachloride and benzene, indicating that at low volumes it arises from a potential energy proportional to $1/V^3$ or $1/R^6$. In the case of the polar derivatives of benzene, the $(T\gamma + B)$ varies approxi-

TABLE 5
 SODIUM CHLORIDE—WATER (10 PER CENT)

	$x_2 = 0.10187$		$x_1 = 0.89813$		$X_2 = 0.0338$			
	P bars	$\frac{10^4}{v_0} \left(\frac{dv}{dT} \right)_P$ deg. ⁻¹	$\frac{10^4}{v_0} \left(\frac{dv}{dP} \right)_T$ bar. ⁻¹	$\left(\frac{dP}{dT} \right)_v$ bars/deg.	$\left(\frac{dE}{dV} \right)_T$ bars	v cc.	V cc.	ϕ_2 cc.
25°	1	3.674	0.3727	9.858	2938	0.93436	18.11	0.3297
	250	3.718	.3485	10.67	2931	.92598	17.95	.3440
	500	3.754	.3272	11.47	2919	.91809	17.79	.3563
	750	3.785	.3082	12.28	2911	.91067	17.65	.3670
	1000	3.812	.2915	13.08	2899	.90367	17.52	.3766
35°	1	4.148	.3702	11.20	3450	.93803	18.18	
	250	4.135	.3464	11.94	3429	.92967	18.02	
	500	4.120	.3253	12.67	3404	.92180	17.87	
	750	4.108	.3066	13.40	3379	.91439	17.72	
	1000	4.093	.2900	14.11	3347	.90740	17.59	
45°	1	4.581	.3708	12.35	3928	.94213	18.26	0.3450
	250	4.515	.3467	13.02	3892	.93372	18.10	.3572
	500	4.454	.3255	13.68	3852	.92581	17.94	.3676
	750	4.397	.3069	14.33	3808	.91837	17.80	.3768
	1000	4.347	.2902	14.98	3765	.91134	17.66	.3848
55°	1	4.988	.3729	13.38	4389	.94665	18.35	
	250	4.871	.3486	13.97	4334	.93815	18.18	
	500	4.763	.3273	14.55	4274	.93016	18.03	
	750	4.667	.3084	15.13	4214	.92264	17.88	
	1000	4.579	.2914	15.71	4154	.91555	17.75	
65°	1	5.378	.3766	14.28	4827	.95157	18.44	0.3502
	250	5.208	.3518	14.80	4754	.94295	18.28	.3619
	500	5.058	.3301	15.32	4680	.93484	18.12	.3720
	750	4.921	.3109	15.83	4602	.92722	17.97	.3809
	1000	4.796	.2937	16.33	4521	.92003	17.83	.3887
75°	1	5.759	.3832	15.03	5231	.95689	18.55	
	250	5.536	.3576	15.48	5139	.94807	18.38	
	500	5.338	.3351	15.93	5045	.93979	18.22	
	750	5.159	.3153	16.36	4945	.93202	18.06	
	1000	5.000	.2977	16.80	4848	.92469	17.92	
85°	1	6.130	.3914	15.66	5607	.96259	18.66	0.3474
	250	5.852	.3647	16.05	5498	.95353	18.48	.3597
	500	5.604	.3415	16.41	5376	.94504	18.32	.3703
	750	5.382	.3209	16.77	5255	.93708	18.16	.3796
	1000	5.185	.3027	17.13	5134	.92958	18.02	.3877

mately as $1/V^{2.7}$. From this result, it is plausible to assume that $(T\gamma + B)$ measures the change of attractive potential energy with volume or the internal pressure due to the attractive forces between the molecules. If such is the case, one might identify B in the Tait equation with the internal pressure due to repulsive forces between the molecules. It should be emphasized, however, that these conclusions are purely tentative and have, as yet, no theoretical support. The quantity $(T\gamma + B)$ is the only function we have

TABLE 6
 SODIUM CHLORIDE—WATER (15 PER CENT)

	$x_2 = 0.14890$		$x_1 = 0.85110$		$X_2 = 0.05116$			
	P bars	$\frac{10^4}{v_0} \left(\frac{dv}{dT} \right)_P$ deg. ⁻¹	$\frac{10^4}{v_0} \left(\frac{dv}{dP} \right)_T$ bar. ⁻¹	$\left(\frac{dP}{dT} \right)_v$ bars/deg.	$\left(\frac{dE}{dV} \right)_T$ bars	v cc.	V cc.	ϕ_2 cc.
25°	1	4.014	0.3401	11.80	3517	.90451	18.17	0.3419
	250	4.005	.3195	12.54	3488	.89709	18.02	.3547
	500	3.997	.3013	13.27	3456	.89007	17.88	.3657
	750	3.988	.2850	13.99	3420	.88345	17.74	.3754
	1000	3.979	.2704	14.72	3388	.87717	17.62	.3839
35°	1	4.351	.3394	12.82	3949	.90829	18.24	
	250	4.303	.3190	13.49	3906	.90085	18.09	
	500	4.259	.3008	14.16	3863	.89381	17.95	
	750	4.217	.2845	14.82	3816	.88718	17.82	
	1000	4.179	.2700	15.48	3769	.88088	17.69	
45°	1	4.677	.3412	13.71	4360	.91241	18.33	0.3555
	250	4.590	.3205	14.32	4305	.90490	18.18	.3664
	500	4.510	.3021	14.93	4249	.89779	18.03	.3758
	750	4.438	.2857	15.53	4190	.89110	17.90	.3841
	1000	4.372	.2710	16.13	4131	.88475	17.77	.3914
55°	1	4.995	.3440	14.52	4763	.91683	18.42	
	250	4.869	.3230	15.07	4694	.90922	18.26	
	500	4.755	.3042	15.63	4628	.90203	18.12	
	750	4.652	.2876	16.18	4559	.89526	17.98	
	1000	4.558	.2728	16.71	4483	.88884	17.85	
65°	1	5.310	.3480	15.26	5158	.92157	18.51	0.3602
	250	5.144	.3265	15.75	5075	.91384	18.35	.3708
	500	4.995	.3074	16.25	4994	.90654	18.21	.3799
	750	4.861	.2904	16.74	4910	.89966	18.07	.3879
	1000	4.739	.2753	17.21	4819	.89314	17.94	.3950
75°	1	5.623	.3543	15.87	5523	.92663	18.61	
	250	5.414	.3320	16.31	5428	.91872	18.45	
	500	5.229	.3123	16.74	5327	.91126	18.30	
	750	5.059	.2948	17.16	5223	.90423	18.16	
	1000	4.906	.2792	17.57	5116	.89759	18.03	
85°	1	5.935	.3620	16.40	5872	.93198	18.72	0.3582
	250	5.680	.3387	16.77	5755	.92386	18.56	.3692
	500	5.454	.3182	17.14	5638	.91620	18.40	.3787
	750	5.251	.3001	17.50	5517	.90901	18.26	.3872
	1000	5.068	.2839	17.85	5392	.90221	18.12	.3944

yet observed which is independent of temperature at constant volume, not only in pure liquids but also in solutions of similar liquids, for example, solutions of aniline in nitrobenzene.¹³ At constant pressure, the expansion coefficients and the compressibilities of normal liquids increase with rise of temperature. At constant volume, the expansion coefficients decrease as the temperature is raised,¹¹ and the same is true for the compressibilities.

Behavior of Water and Glycol. The well-known exceptional behavior of

TABLE 7
 SODIUM CHLORIDE—WATER (20 PER CENT)

	$x_2 = 0.19957$		$x_1 = 0.80043$		$X_2 = 0.07136$			
	P bars	$\frac{10^4}{v_0} \left(\frac{dv}{dT} \right)_P$ deg ⁻¹	$\frac{10^4}{v_0} \left(\frac{dv}{dP} \right)_T$ bars ⁻¹	$\left(\frac{dP}{dT} \right)_V$ bars/deg.	$\left(\frac{dE}{dV} \right)_T$ bars.	v cc.	V cc.	ϕ_2 cc.
25°	1	4.264	0.3083	13.83	4122	0.87329	18.25	0.3533
	250	4.213	.2910	14.48	4066	.86677	18.12	.3645
	500	4.164	.2755	15.11	4004	.86059	17.99	.3744
	750	4.120	.2617	15.74	3942	.85474	17.87	.3831
	1000	4.079	.2492	16.37	3880	.84916	17.75	.3907
35°	1	4.510	.3091	14.59	4494	.87713	18.33	
	250	4.431	.2917	15.19	4430	.87057	18.20	
	500	4.360	.2761	15.79	4365	.86434	18.07	
	750	4.295	.2622	16.38	4297	.85845	17.94	
	1000	4.234	.2497	16.96	4225	.85284	17.83	
45°	1	4.756	.3123	15.23	4844	.88120	18.42	0.3652
	250	4.649	.2946	15.78	4770	.87454	18.28	.3749
	500	4.553	.2787	16.34	4698	.86823	18.15	.3834
	750	4.466	.2645	16.88	4620	.86225	18.02	.3908
	1000	4.385	.2517	17.42	4541	.85657	17.90	.3974
55°	1	4.999	.3156	15.84	5196	.88551	18.51	
	250	4.865	.2976	16.35	5114	.87875	18.37	
	500	4.744	.2813	16.86	5032	.87244	18.24	
	750	4.633	.2669	17.36	4946	.86629	18.11	
	1000	4.533	.2539	17.85	4857	.86052	17.99	
65°	1	5.243	.3199	16.39	5540	.89006	18.60	0.3697
	250	5.081	.3012	16.87	5454	.88318	18.46	.3792
	500	4.934	.2847	17.33	5359	.87666	18.32	.3873
	750	4.801	.2699	17.79	5265	.87050	18.20	.3944
	1000	4.678	.2565	18.24	5167	.86465	18.07	.4009
75°	1	5.486	.3259	16.83	5858	.89485	18.70	
	250	5.291	.3066	17.26	5758	.88781	18.56	
	500	5.116	.2894	17.68	5654	.88114	18.42	
	750	4.957	.2741	18.08	5544	.87485	18.29	
	1000	4.813	.2604	18.48	5433	.86887	18.16	
85°	1	5.726	.3330	17.20	6158	.89988	18.81	0.3685
	250	5.497	.3129	17.57	6042	.89265	18.66	.3783
	500	5.292	.2950	17.94	5924	.88581	18.52	.3868
	750	5.107	.2791	18.30	5803	.87936	18.38	.3942
	1000	4.937	.2649	18.64	5675	.87324	18.25	.4007

water is emphasized by a consideration of the $(\partial E / \partial V)_T - V$ diagram for this liquid, which is shown in FIGURE 5. It will be seen that the effect of temperature at constant volume on $(\partial E / \partial V)_T$ is enormous and has the opposite sign from that observed in other liquids. The dotted curves in FIGURE 5 represent the results for aniline on the same scale and give a comparison of the behavior of water with that of a fairly regular liquid. It will be seen that C_V for water decreases as the liquid is compressed. In FIGURE 6, we

TABLE 8
SODIUM CHLORIDE—WATER (25 PER CENT)

	$x_2 = 0.24929$			$x_1 = 0.75071$		$X_2 = 0.09284$		
	P bars	$\frac{10^4}{v_0} \left(\frac{dv}{dT} \right)_P$ deg. ⁻¹	$\frac{10^4}{v_0} \left(\frac{dv}{dP} \right)_T$ bars ⁻¹	$\left(\frac{dP}{dT} \right)_V$ bars/deg.	$\left(\frac{dE}{dV} \right)_T$ cc.	v cc.	V cc.	ϕ_2 cc.
25°	1	4.471	0.2812	15.90	4739	0.84348	18.36	0.3633
	250	4.388	.2665	16.47	4660	.83773	18.24	.3732
	500	4.312	.2534	17.02	4574	.83225	18.12	.3819
	750	4.242	.2414	17.57	4488	.82705	18.01	.3895
	1000	4.178	.2306	18.12	4402	.82208	17.90	.3962
35°	1	4.637	.2828	16.40	5052	.84732	18.45	
	250	4.539	.2680	16.94	4969	.84151	18.32	
	500	4.450	.2546	17.48	4886	.83597	18.20	
	750	4.369	.2425	18.02	4802	.83071	18.09	
	1000	4.294	.2316	18.54	4712	.82570	17.98	
45°	1	4.812	.2872	16.75	5327	.85134	18.53	0.3740
	250	4.697	.2719	17.27	5244	.84542	18.41	.3826
	500	4.590	.2582	17.78	5156	.83977	18.28	.3900
	750	4.495	.2457	18.29	5068	.83442	18.17	.3966
	1000	4.406	.2345	18.79	4977	.82931	18.05	.4025
55°	1	4.997	.2905	17.20	5642	.85552	18.63	
	250	4.861	.2748	17.69	5554	.84951	18.49	
	500	4.739	.2607	18.18	5465	.84377	18.37	
	750	4.626	.2480	18.65	5369	.83834	18.25	
	1000	4.524	.2366	19.12	5273	.83316	18.14	
65°	1	5.186	.2953	17.56	5936	.85989	18.72	0.3784
	250	5.027	.2791	18.01	5839	.85375	18.59	.3867
	500	4.882	.2646	18.45	5738	.84790	18.46	.3939
	750	4.751	.2515	18.89	5637	.84236	18.34	.4002
	1000	4.630	.2397	19.32	5532	.83708	18.22	.4058
75°	1	5.378	.3005	17.90	6230	.86445	18.82	
	250	5.191	.2838	18.29	6117	.85817	18.68	
	500	5.022	.2687	18.69	6006	.85219	18.55	
	750	4.869	.2553	19.07	5888	.84654	18.43	
	1000	4.728	.2432	19.44	5767	.84115	18.31	
85°	1	5.573	.3080	18.09	6477	.86919	18.92	0.3778
	250	5.353	.2904	18.43	6350	.86272	18.78	.3864
	500	5.153	.2746	18.77	6222	.85658	18.65	.3939
	750	4.972	.2606	19.08	6083	.85077	18.52	.4005
	1000	4.808	.2479	19.39	5944	.84525	18.40	.4062

have illustrated the energy-volume coefficient of ethylene glycol as a function of volume at different temperatures and pressures. This liquid behaves qualitatively like water, but the magnitude of the temperature effect is very much less. In FIGURE 7, $(T\gamma + B)$ is plotted for water and for ethylene glycol. In both cases, it is not a pure volume function; indeed, in water, the variation of $(T\gamma + B)$ with temperature at constant volume is still large, although not so striking as the variation in $(\partial E / \partial V)_T$ itself.

TABLE 9
 SODIUM BROMIDE—WATER (5 PER CENT)

	$x_2 = 0.0491$		$x_1 = 0.9509$		$X_2 = 0.0090$			
	P bars	$\frac{10^4}{v_0} \left(\frac{dv}{dT} \right)_P$ deg. ⁻¹	$\frac{10^4}{v_0} \left(\frac{dv}{dP} \right)_T$ bars ⁻¹	$\left(\frac{dP}{dT} \right)_V$ bars/deg.	$\left(\frac{dE}{dV} \right)_T$ bars	v cc.	V cc.	ϕ_2 cc.
25°	1	3.038	0.4333	7.011	2089	0.96550	18.13	0.2400
	250	3.217	.4014	8.014	2139	.95547	17.94	.2480
	500	3.370	.3738	9.016	2188	.94612	17.76	.2547
	750	3.502	.3498	10.01	2234	.93740	17.60	.2606
	1000	3.616	.3285	11.01	2282	.92921	17.45	.2657
35°	1	3.733	0.4267	8.749	2695	.96877	18.19	
	250	3.798	.3957	9.598	2707	.95886	18.00	
	500	3.850	.3688	10.44	2717	.94961	17.83	
	750	3.896	.3454	11.28	2725	.94097	17.67	
	1000	3.934	.3247	12.12	2734	.93286	17.52	
45°	1	4.383	0.4246	10.32	3282	.97272	18.26	0.2533
	250	4.353	.3938	11.05	3265	.96282	18.08	.2598
	500	4.323	.3672	11.77	3244	.95357	17.90	.2649
	750	4.296	.3440	12.49	3223	.94493	17.74	.2694
	1000	4.269	.3234	13.20	3199	.93682	17.59	.2734
55°	1	4.985	.4254	11.72	3844	.97730	18.35	
	250	4.875	.3945	12.36	3805	.96731	18.16	
	500	4.777	.3677	12.99	3762	.95803	17.99	
	750	4.687	.3444	13.61	3715	.94933	17.83	
	1000	4.607	.3239	14.22	3666	.94117	17.67	
65°	1	5.542	.4294	12.91	4364	.98245	18.45	0.2590
	250	5.363	.3981	13.47	4304	.97234	18.26	.2649
	500	5.205	.3709	14.03	4244	.96290	18.08	.2698
	750	5.064	.3471	14.59	4183	.95409	17.91	.2740
	1000	4.937	.3262	15.13	4115	.94583	17.76	.2781
75°	1	6.053	.4362	13.88	4831	.98817	18.55	
	250	5.815	.4038	14.40	4763	.97785	18.36	
	500	5.605	.3758	14.91	4690	.96823	18.18	
	750	5.420	.3515	15.42	4618	.95925	18.01	
	1000	5.253	.3301	15.91	4538	.95084	17.85	
85°	1	6.517	.4457	14.62	5234	.99441	18.67	0.2590
	250	6.229	.4119	15.12	5164	.98380	18.47	.2653
	500	5.976	.3828	15.61	5090	.97393	18.29	.2706
	750	5.753	.3576	16.09	5012	.96473	18.11	.2755
	1000	5.553	.3355	18.55	4927	.95613	17.95	.2795

It has been shown¹¹ that the very anomalous behavior of the $(\partial E / \partial V)_T$ vs V curves for water shown in FIGURE 5 may be accounted for qualitatively in all aspects in terms of the current theories of the molecular distribution in this liquid.* The low coordination number (approximately 4) for water, and other evidence, indicate that weak directed bonds between the water molecules tend to hold them in an open structure. When the liquid is compressed,

TABLE 10
 SODIUM BROMIDE—WATER (15 PER CENT)

	$x_2 = 0.14997$		$x_1 = 0.85003$		$X_2 = 0.0300$			
	P bars	$\frac{10^4}{v_0} \left(\frac{dv}{dT} \right)_P$ deg. ⁻¹	$\frac{10^4}{v_0} \left(\frac{dv}{dP} \right)_T$ bars ⁻¹	$\left(\frac{dP}{dT} \right)_V$ bars/deg.	$\left(\frac{dE}{dV} \right)_T$ bars	v cc.	V cc.	ϕ_2 cc.
25°	1	3.805	0.3873	9.824	2928	0.88999	18.30	0.2501
	250	3.836	.3611	10.62	2916	.88170	18.13	.2569
	500	3.864	.3381	11.43	2907	.87393	17.97	.2627
	750	3.888	.3180	12.23	2896	.86664	17.82	.2679
	1000	3.906	.3000	13.02	2881	.85976	17.68	.2723
35°	1	4.267	.3845	11.10	3419	.89359	18.37	
	250	4.245	.3587	11.83	3395	.88532	18.20	
	500	4.221	.3359	12.57	3373	.87757	18.04	
	750	4.200	.3160	13.29	3345	.87030	17.89	
	1000	4.181	.2983	14.02	3320	.86344	17.75	
45°	1	4.710	.3845	12.25	3896	.89760	18.46	0.2617
	250	4.633	.3586	12.92	3860	.88930	18.29	.2672
	500	4.565	.3359	13.59	3823	.88151	18.13	.2719
	750	4.502	.3160	14.25	3783	.87421	17.98	.2760
	1000	4.444	.2982	14.90	3740	.86732	17.83	.2796
55°	1	5.133	.3868	13.27	4353	.90205	18.55	
	250	5.003	.3606	13.87	4301	.89366	18.38	
	500	4.887	.3377	14.47	4248	.88580	18.21	
	750	4.782	.3175	15.06	4191	.87841	18.06	
	1000	4.687	.2996	15.64	4131	.87146	17.92	
65°	1	5.538	.3916	14.14	4780	.90687	18.65	0.2671
	250	5.354	.3648	14.68	4713	.89833	18.47	.2722
	500	5.189	.3413	15.20	4639	.89033	18.31	.2765
	750	5.041	.3207	15.72	4565	.88284	18.15	.2803
	1000	4.909	.3025	16.23	4487	.87578	18.01	.2837
75°	1	5.922	.3979	14.88	5179	.91208	18.75	
	250	5.679	.3701	15.34	5090	.90337	18.57	
	500	5.464	.3460	15.79	4996	.89520	18.41	
	750	5.274	.3249	16.23	4900	.88757	18.25	
	1000	5.103	.3061	16.67	4803	.88038	18.10	
85°	1	6.286	.4060	15.48	5542	.91767	18.87	0.2679
	250	5.981	.3773	15.85	5426	.90873	18.68	.2733
	500	5.713	.3522	16.22	5308	.90036	18.51	.2779
	750	5.476	.3303	16.58	5187	.89255	18.35	.2821
	1000	5.262	.3109	16.93	5063	.88519	18.20	.2855

these bonds are broken, molecules moving into the empty spaces, and a corresponding increase in potential energy results. The contribution of this effect to $(\partial E/\partial V)_T$ is negative, it has the same sign as the internal pressure arising from the repulsive forces. The magnitude of the contribution to $(\partial E/\partial V)_T$ arising from the breaking of hydrogen bonds during the compression of water may be estimated roughly by plotting $(\partial E/\partial V)_T$ against temperature at constant volume and determining by extrapolation the value of

TABLE 11
 SODIUM BROMIDE—WATER (25 PER CENT)

	$x_2 = 0.25004$		$x_2 = 0.74996$		$X_2 = 0.05514$			
	P bars	$\frac{10^4}{v_0} \left(\frac{dv}{dT} \right)_P$ deg. ⁻¹	$\frac{10^4}{v_0} \left(\frac{dv}{dP} \right)_T$ bars ⁻¹	$\left(\frac{dP}{dT} \right)_V$ bars/deg.	$\left(\frac{dE}{dT} \right)_T$ bars	v cc.	V cc.	ϕ_2 cc.
25°	1	4.408	0.3424	12.87	3836	0.81661	18.54	0.2580
	250	4.388	.3213	13.66	3822	.80986	18.38	.2638
	500	4.367	.3026	14.43	3802	.80351	18.24	.2689
	750	4.348	.2859	15.21	3784	.79750	18.10	.2733
	1000	4.331	.2711	15.98	3764	.79182	17.97	.2772
35°	1	4.696	.3428	13.70	4220	.82034	18.62	
	250	4.627	.3216	14.39	4184	.81356	18.47	
	500	4.565	.3028	15.08	4146	.80717	18.32	
	750	4.509	.2801	15.76	4106	.80113	18.19	
	1000	4.458	.2712	16.44	4065	.79542	18.06	
45°	1	4.977	.3448	14.43	4589	.82432	18.71	0.2680
	250	4.867	.3233	15.05	4537	.81746	18.56	.2728
	500	4.770	.3044	15.67	4485	.81101	18.41	.2790
	750	4.679	.2875	16.27	4425	.80491	18.27	.2806
	1000	4.598	.2725	16.87	4366	.79914	18.14	.2837
55°	1	5.255	.3486	15.07	4943	.82854	18.81	
	250	5.109	.3266	15.64	4881	.82158	18.65	
	500	4.979	.3073	16.20	4815	.81502	18.50	
	750	4.859	.2900	16.76	4749	.80884	18.36	
	1000	4.752	.2747	17.30	4676	.80299	18.23	
65°	1	5.527	.3535	15.64	5287	.83302	18.91	0.2730
	250	5.348	.3310	16.16	5214	.82592	18.75	.2774
	500	5.190	.3110	16.69	5143	.81925	18.60	.2813
	750	5.045	.2935	17.19	5062	.81296	18.45	.2846
	1000	4.916	.2778	17.70	4984	.80701	18.32	.2875
75°	1	5.794	.3595	16.12	5610	.83776	19.02	
	250	5.588	.3363	16.62	5535	.83050	18.85	
	500	5.404	.3157	17.12	5459	.82369	18.70	
	750	5.239	.2976	17.60	5377	.81727	18.55	
	1000	5.089	.2815	18.08	5294	.81121	18.41	
85°	1	6.055	.3668	16.51	5911	.84273	19.13	0.2742
	250	5.826	.3426	17.01	5841	.83529	18.96	.2788
	500	5.622	.3212	17.50	5767	.82831	18.80	.2829
	750	5.439	.3025	17.98	5689	.82174	18.65	.2864
	1000	5.275	.2859	18.45	5607	.81554	18.51	.2893

$(\partial E/\partial V)_T$ where the temperature effect becomes zero. This extrapolation¹¹ along the line for $V = 18.12$ cc indicates that $(\partial E/\partial V)_T$ for water becomes independent of temperature and has a value of 6,300 bars in the vicinity of 200°C. The difference between this value and the observed value of $(\partial E/\partial V)_T$ at any lower pressure is taken as a measure of the contribution of the breaking of the hydrogen bonds to $(\partial E/\partial V)_T$ for water. Calculations show that the breaking of a few per cent of the hydrogen bonds in the liquid

TABLE 12
SODIUM BROMIDE—WATER (35 PER CENT)

	$x_2 = 0.34554$		$x_1 = 0.65446$		$X_2 = 0.08460$			
	P bars	$\frac{10^4}{v_0} \left(\frac{dv}{dT} \right)_P$ deg. ⁻¹	$\frac{10^4}{v_0} \left(\frac{dv}{dP} \right)_T$ bars ⁻¹	$\left(\frac{dP}{dT} \right)_v$ bars/deg.	$\left(\frac{dE}{dV} \right)_T$ bars	v cc.	V cc.	ϕ_2 cc.
25°	1	4.841	0.3006	16.10	4798	0.74780	18.85	0.2647
	250	4.749	.2838	16.73	4737	.74236	18.71	.2697
	500	4.666	.2687	17.37	4678	.73720	18.58	.2740
	750	4.589	.2551	17.99	4613	.73230	18.46	.2778
	1000	4.521	.2430	18.60	4545	.72765	18.34	.2811
35°	1	5.002	.3034	16.49	5080	.75150	18.94	
	250	4.885	.2862	17.07	5009	.74598	18.80	
	500	4.784	.2709	17.66	4941	.74075	18.67	
	750	4.689	.2572	18.23	4867	.73579	18.54	
	1000	4.604	.2447	18.81	4795	.73108	18.42	
45°	1	5.169	.3072	16.83	5353	.75532	19.04	0.2734
	250	5.032	.2895	17.38	5279	.74971	18.89	.2775
	500	4.909	.2738	17.93	5204	.74440	18.76	.2810
	750	4.797	.2598	18.46	5122	.73936	18.63	.2841
	1000	4.696	.2470	19.01	5047	.73458	18.51	.2868
55°	1	5.339	.3119	17.12	5616	.75930	19.14	
	250	5.182	.2937	17.64	5538	.75357	18.99	
	500	5.041	.2775	18.17	5462	.74816	18.85	
	750	4.912	.2630	18.68	5379	.74302	18.73	
	1000	4.795	.2501	19.17	5290	.73816	18.60	
65°	1	5.515	.3171	17.39	5879	.76344	19.24	0.2780
	250	5.336	.2983	17.89	5799	.75759	19.09	.2818
	500	5.178	.2816	18.39	5718	.75206	18.95	.2851
	750	5.034	.2667	18.88	5633	.74683	18.82	.2879
	1000	4.903	.2533	19.36	5546	.74187	18.70	.2904
75°	1	5.695	.3232	17.62	6133	.76773	19.35	
	250	5.498	.3036	18.11	6054	.76174	19.20	
	500	5.322	.2863	18.59	5971	.75608	19.05	
	750	5.162	.2709	19.06	5885	.75073	18.92	
	1000	5.019	.2571	19.52	5795	.74565	18.79	
85°	1	5.879	.3301	17.81	6377	.77218	19.46	0.2795
	250	5.664	.3098	18.28	6296	.76603	19.31	.2834
	500	5.472	.2918	18.75	6214	.76023	19.16	.2869
	750	5.298	.2757	19.22	6133	.75475	19.02	.2898
	1000	5.140	.2615	19.66	6040	.74957	18.89	.2923

by the application of a pressure of 1,000 bars accounted for the anomalous effect of temperature on the energy-volume coefficient of water.

Solutions. The foregoing results do not give much promise that the thermodynamic properties of solutions of electrolytes dissolved in water will exhibit many simple regularities at low temperatures. They do, however, indicate that solutions in ethylene glycol will exhibit less complex behavior and, therefore, that such solutions are suitable to act as models in any dis-

TABLE 13
 SODIUM BROMIDE—WATER (45 PER CENT)

	$x_2 = 0.44338$		$x_1 = 0.55662$		$X_2 = 0.12238$			
	P bars	$\frac{10^4}{v_0} \left(\frac{dv}{dT} \right)_P$ deg. ⁻¹	$\frac{10^4}{v_0} \left(\frac{dv}{dT} \right)_T$ bars ⁻¹	$\left(\frac{dP}{dT} \right)_V$ bars/deg.	$\left(\frac{dE}{dV} \right)_T$ bars	v cc.	V cc.	ϕ_2 cc.
25°	1	5.141	0.2614	19.67	5863	0.67832	19.27	0.2710
	250	5.011	.2483	20.18	5766	.67402	19.15	.2750
	500	4.893	.2363	20.71	5674	.66991	19.03	.2785
	750	4.788	.2254	21.24	5582	.66600	18.92	.2816
	1000	4.690	.2157	21.74	5481	.66226	18.81	.2843
35°	1	5.223	.2656	19.66	6056	.68184	19.37	
	250	5.083	.2518	20.19	5971	.67745	19.25	
	500	4.959	.2395	20.71	5881	.67326	19.13	
	750	4.844	.2284	21.21	5785	.66927	19.01	
	1000	4.739	.2182	21.72	5692	.66547	18.91	
45°	1	5.311	.2702	19.66	6253	.68544	19.47	0.2784
	250	5.163	.2560	20.17	6166	.68095	19.35	.2817
	500	5.027	.2432	20.67	6075	.67667	19.22	.2846
	750	4.903	.2317	21.16	5981	.67261	19.11	.2871
	1000	4.789	.2213	21.64	5884	.66873	19.00	.2894
55°	1	5.407	.2759	19.60	6430	.68912	19.58	
	250	5.243	.2611	20.08	6338	.68452	19.45	
	500	5.096	.2477	20.57	6249	.68013	19.32	
	750	4.961	.2358	21.04	6153	.67597	19.20	
	1000	4.838	.2249	21.51	6057	.67200	19.09	
65°	1	5.510	.2808	19.62	6633	.69290	19.69	0.2827
	250	5.333	.2654	20.09	6542	.68819	19.55	.2858
	500	5.171	.2517	20.54	6445	.68371	19.42	.2885
	750	5.024	.2393	20.99	6347	.67946	19.30	.2908
	1000	4.891	.2282	21.43	6245	.67542	19.19	.2929
75°	1	5.622	.2875	19.55	6804	.69677	19.80	
	250	5.425	.2714	19.99	6709	.69193	19.66	
	500	5.247	.2569	20.42	6608	.68732	19.53	
	750	5.086	.2440	20.84	6504	.68296	19.40	
	1000	4.940	.2324	21.26	6401	.67882	19.29	
85°	1	5.740	.2945	19.49	6978	.70073	19.91	0.2846
	250	5.521	.2777	19.88	6869	.69574	19.77	.2877
	500	5.323	.2626	20.27	6759	.69101	19.63	.2904
	750	5.145	.2492	20.65	6645	.68653	19.50	.2928
	1000	4.983	.2370	21.03	6531	.68228	19.38	.2948

cussion of the behavior of water solutions. With this in mind, we have examined the behavior of sodium bromide both in water and glycol and will compare the results later.

In general, the volume change produced by the solution of an electrolyte in a solvent would be expected to be determined by the long range electrostatic ion-molecule forces and the molecule-molecule interactions in the liquid. It has been known for a long time that the compressibility of aque-

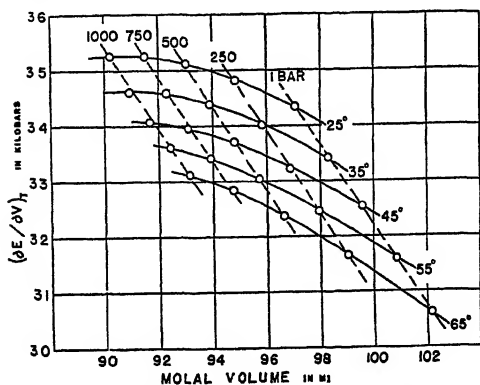


FIGURE 1. The energy-volume coefficients of carbon tetrachloride as a function of volume at different temperatures.

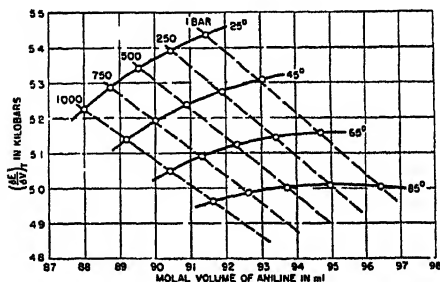
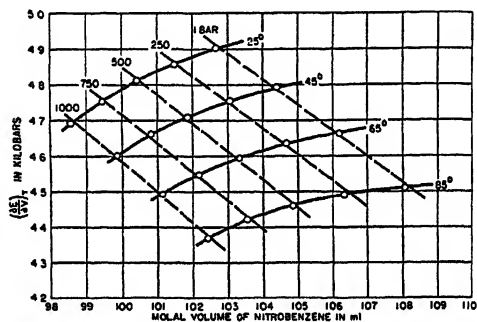


FIGURE 2. The energy-volume coefficients of nitrobenzene and aniline as functions of volume at different temperatures.

ous solutions may be related to the volume change on mixing by the simple Tammann hypothesis that water in a solution behaves as water under a given hydrostatic pressure. It is possible to make quantitative use of this relationship, and, indeed, the compressibility of solutions may be computed from apparent volumes and the compressibility of pure water.¹⁴ It must be emphasized, however, that this relation holds only at constant temperature

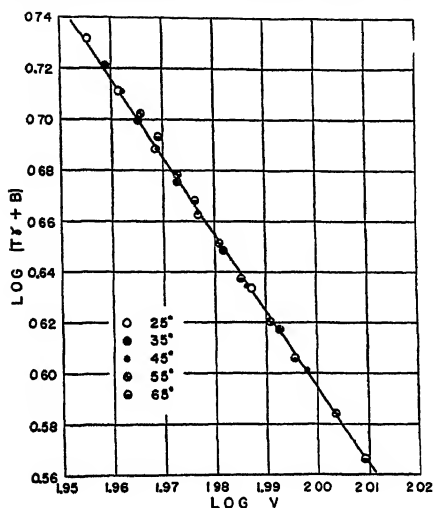


FIGURE 3 Illustration of $(T\gamma + B)$ as a pure volume function for carbon tetrachloride.

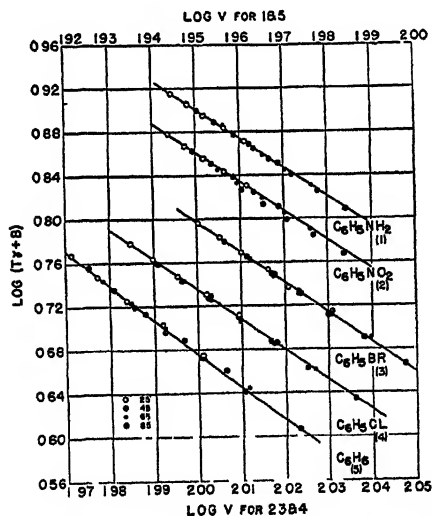


FIGURE 4. $\text{Log } (T\gamma + B)$ as a function of $\text{log } V$ for benzene and some of its derivatives. (Note: $T\gamma + B$ is a pure volume function for these liquids.)

and that the behavior of water in a solution under changes of temperature cannot be predicted from a knowledge of the behavior of water under pressure. It will also be recalled that the apparent volumes of salts *increase* with temperature up to about 50° C., even though any simple picture of internal forces would lead one to expect that the temperature coefficient of this quantity would be negative. Furthermore, simple electrostatic consid-

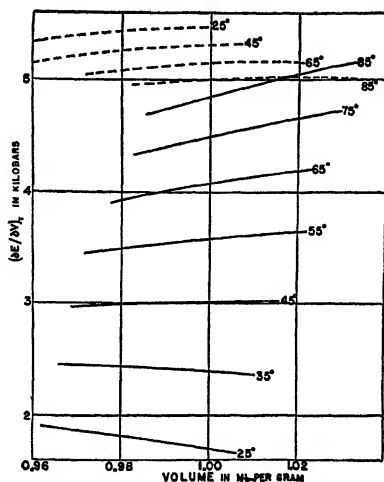


FIGURE 5. The energy-volume coefficient of water as a function of volume at different temperatures. (The dotted lines compare the behavior of water with that of aniline.)

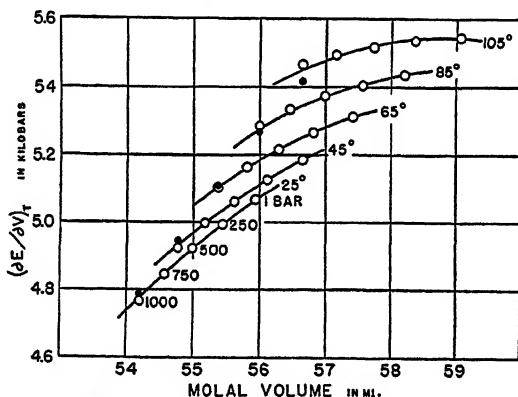


FIGURE 6. Energy-volume coefficient of ethylene glycol as a function of volume at different temperatures.

erations would lead to the conclusion that the volume changes in infinite dilution produced by monovalent salts should be approximately proportional to the reciprocal of the mean ionic radii according to the relation

$$(V_2^{\infty} - V_2^i) = \frac{-Nq^2}{2a} \frac{1}{K^2} \frac{dK}{dP} \quad (7)$$

where $(V_2^{\infty} - V_2^i)$ is the contraction when 1 mol of salt is dissolved to give an infinitely dilute solution, q is the valence of the ions, a is the mean ionic radius, and K is the dielectric constant of the medium. This relation has been explored by Kritschewsky¹⁵ who found that it described well the effect of pressure on the partial molal volumes of salts in water. However, a study of his paper indicates that the values of the ionic radii obtained from standard

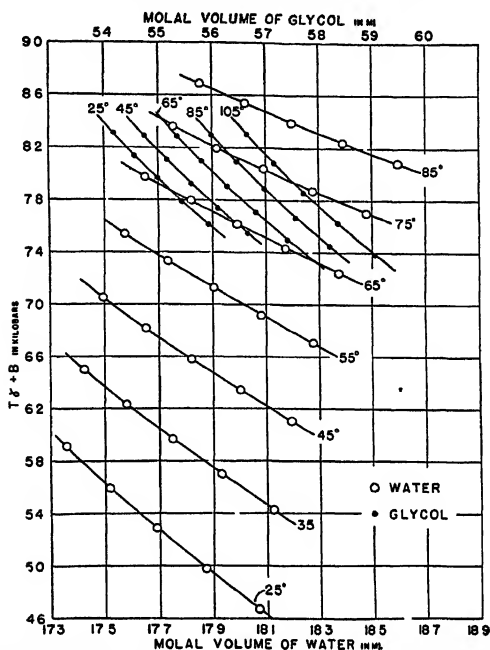


FIGURE 7. The variation of $(T\gamma + B)$ with volume at different temperatures for water and glycol. (Note the absence of any indication of a pure volume function.)

sources are by no means in agreement with those required by the application of EQUATION 7. In FIGURE 8, we have plotted the contractions of salts on solution in water against the reciprocal of the ionic radius taken from crystal data. It will be seen that the relation does not hold for the alkali bromides in water. On the other hand, the relationship holds qualitatively quite well

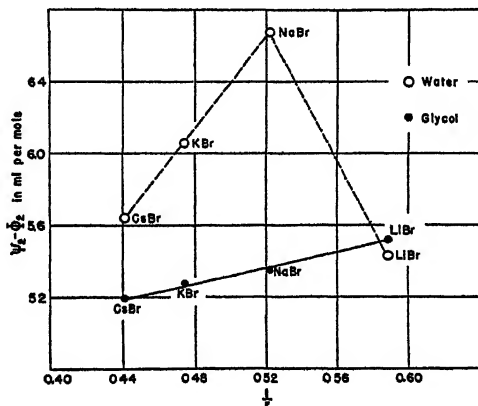


FIGURE 8. Contraction on mixing as a function of reciprocal of mean ionic radius for different salts in water and in glycol.

for solutions in ethylene glycol, as may be seen from the lower curve, and, indeed, is an example of a regularity of behavior of solutions of electrolytes in this solvent. It is, therefore, desirable to recall some of these regularities.

Solutions in Ethylene Glycol. FIGURE 9 shows that the apparent volumes

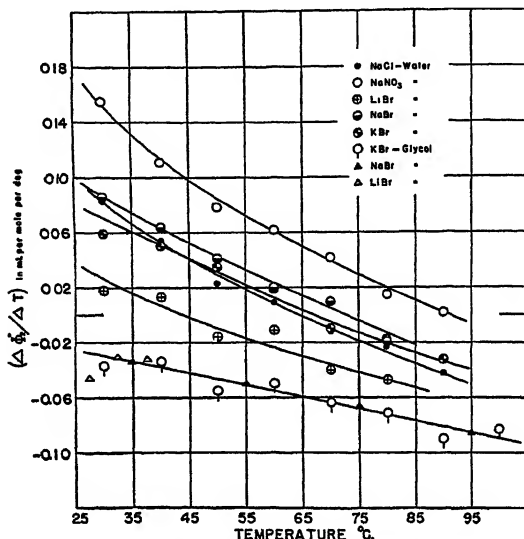


FIGURE 9. The mean apparent expansion coefficient of different salts in water and in glycol as a function of temperature

of various salts dissolved in ethylene glycol decrease with temperature. The apparent expansion coefficients are negative as contrasted with those for aqueous solutions which are all positive at the lower temperatures. The compressibility of ethylene glycol solutions is, of course, determined by the value of B in the Tait equation for the solutions, and the Tammann hypothesis may be applied quantitatively to compute compressibilities from volume changes on solution and the properties of pure glycol. FIGURE 10 shows that the quantity B also determines the expansion coefficients of these solutions at different pressures and temperatures. Furthermore, the apparent volume itself depends only on the specific volume of the solution and is independent of temperature as long as the total volume is kept constant. This is strikingly illustrated in FIGURE 11, where the apparent volumes of sodium bromide in ethylene glycol are plotted against volume, and for each solution the points observed at temperatures between 25° and 105° all fall on the same curve.

The energy-volume coefficients of glycol-sodium bromide solutions are shown as functions of the volume in FIGURE 12. In general, curves for the solutions resemble those for the pure solvent, but it should be remarked that, in the more concentrated solutions, the variation with temperature at con-

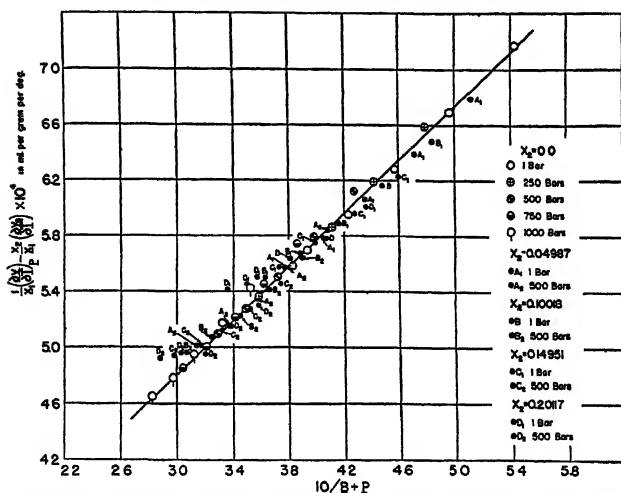


FIGURE 10. Expansion coefficients of glycol and sodium bromide solutions in glycol as functions of $(B' + P)$, the constant computed from the compressibility and the Tait equation.

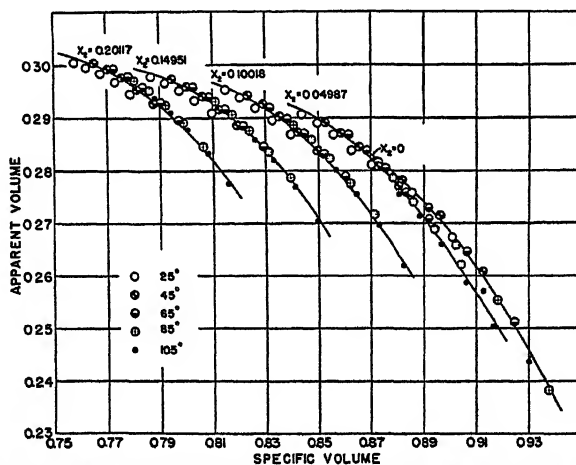


FIGURE 11. The apparent volumes of sodium bromide in solutions of glycol of different concentrations as a function of the specific volume of the solution at different temperatures.

stant volume has changed sign and now exhibits the type of behavior found with carbon tetrachloride.

The effect of addition of salt on the energy-volume coefficient of glycol is given in FIGURE 13, which shows that the addition of salt causes very little variation in $(\partial E/\partial V)_T$ for the system, a result which would be expected from the fact that the apparent volume depends only on the total volume. It seems, therefore, that the contraction which takes place when salt is added to glycol is just sufficient to produce a balance between the extra attractive component of the internal pressure due to the ion-molecule forces and the

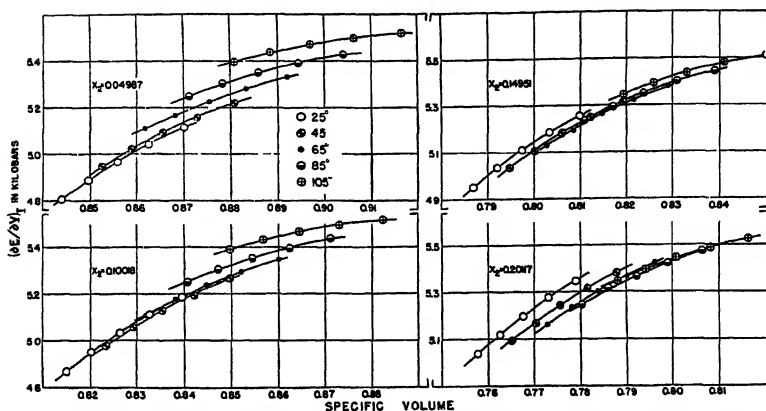


FIGURE 12. Energy-volume coefficient of solutions of sodium bromide in glycol as functions of the volume.

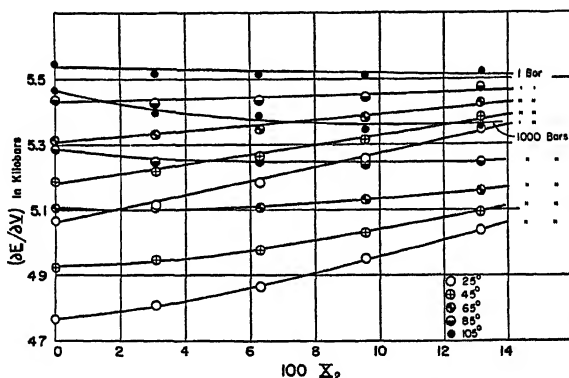


FIGURE 13. The effect of concentration on the energy-volume coefficient of solutions of sodium bromide in glycol.

repulsive component of the internal pressure, so that the total quantity remains constant.

We may summarize the behavior of glycol solutions by stating that the volume changes on mixing, the compressibilities and the energy-volume coefficients all behaving as if the interaction between the dissolved ions and the solvent merely causes a tightening up of the structure, such as would be produced by stronger forces between the ions and the molecules than between the molecules themselves. Furthermore, the main effects are due to these long-range forces, and the effects of short-range forces or of changes in the molecular distribution in glycol are only secondary.

Water Solutions. In discussing the results for aqueous solutions, we must limit ourselves to a survey of the general features and postpone a quantitative discussion of the results. This survey offers promise that quantitative relations among the various properties may be established at least on a semi-

empirical basis, but the complexity of the phenomena has prevented us from carrying out this part of the work. A sample of the type of results found for aqueous solutions is given in FIGURE 14, where the energy-volume coefficients

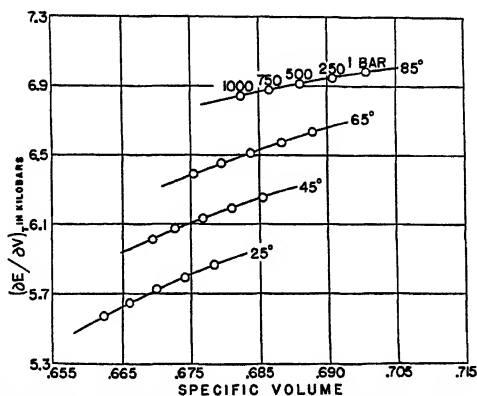


FIGURE 14. The energy-volume coefficient of a 44% solution of sodium bromide in water as a function of the specific volume at different temperatures.

for a 44 per cent solution of sodium bromide are plotted against volume. This figure may be contrasted with FIGURE 12. It is interesting to note that the curves are much closer together than on the diagram for pure water, but are still widely separated when compared with the glycol solutions.

In contrast to the behavior observed in glycol solutions, the expansion coefficients are by no means determined by $(B + P)$, although it will be recalled that at any constant temperature the apparent volumes and the compressibilities may be related through the Tammann hypothesis. In FIGURE 15, it will be seen that the expansion coefficient is a different function

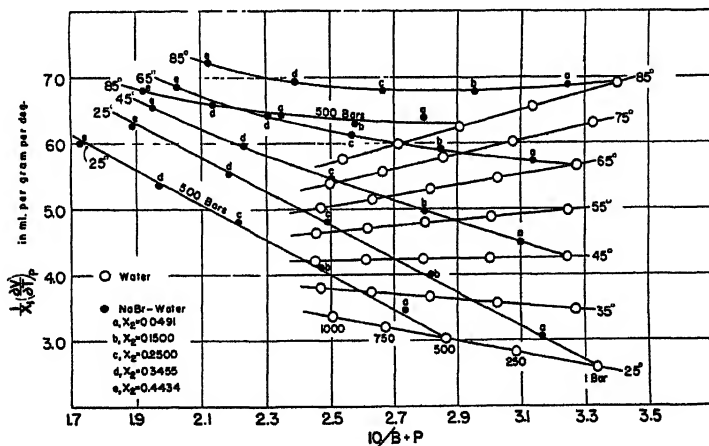


FIGURE 15. Expansion coefficients of solutions of sodium bromide in water as a function of the constant B in the Tait equation.

of $(B + P)$ for each temperature, although there is an indication that at higher temperatures the curves may come together and give a behavior similar to that observed in glycol solutions.

In FIGURE 16, the effect of concentration on the constant B in the Tait

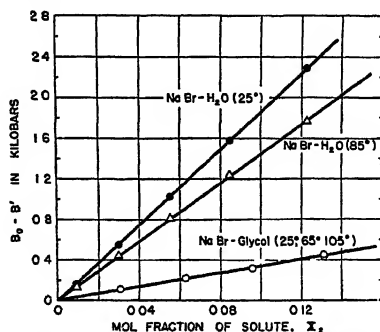


FIGURE 16. The variation of the constant B in the Tait equation with concentration for solutions of sodium bromide in water and in ethylene glycol.

equation is compared for sodium bromide in glycol and in water. The quantity plotted is $(B_0 - B')$, B' being the value of B obtained from the compressibility results for the solutions through EQUATION 2. $(B_0 - B')$ measures the effect of salt on the compressibility of the system. It is interesting to note that the variation of B' with concentration in glycol solutions is independent of temperature and is much smaller than that encountered in aqueous solutions, where a marked temperature dependence of $(B_0 - B')$ at any given concentration is also observed.

Similar behavior is noticed in the effects of concentration on the energy-volume coefficients in aqueous solutions as compared with the same results for glycol solutions. These results are illustrated in FIGURE 17, where fea-

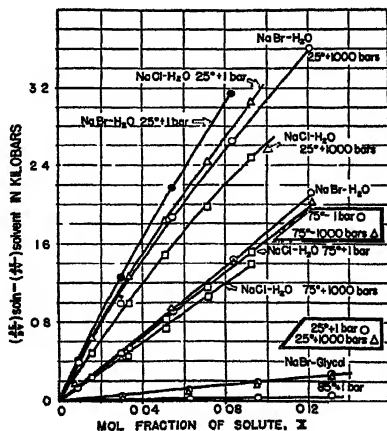


FIGURE 17. The effect of concentration on the energy-volume coefficients of different salts in water and in ethylene glycol at different temperatures and pressures.

tures of interest are, first, the small variation of $(\partial E/\partial V)_T$ with concentration in glycol solutions as compared with the relatively large variation in aqueous solutions, and, second, the marked effects of temperature in aqueous solutions. It will be noted that sodium chloride and sodium bromide form a fairly compact family of curves at 25° both at 1 and 1,000 bars pressure. There is, however, enough divergence among these curves to indicate significant differences. At 75°, another compact family is observed, the effect of salt is much reduced and indeed, is almost the same for the two salts at both pressures. FIGURE 17 points to the conclusions that, at higher temperatures, the effect of concentration on $(\partial E/\partial V)_T$ in aqueous solutions probably approaches the value observed in glycol solutions, that specific salt effects are changed as the temperature is raised, and that at higher temperatures the solution adjusts itself in volume upon addition of salt in such a way that $(\partial E/\partial V)_T$ remains substantially constant. It is interesting to note that unpublished results for aqueous solutions of lithium bromide indicate a very small effect of concentration on $(\partial E/\partial V)_T$ at 25° and that at 85° this salt actually causes a lowering of the energy-volume coefficient. This is undoubtedly connected with the small volume change on mixing shown in the system lithium bromide-water, but a complete discussion must be reserved for a later paper.

In water solutions, the apparent volume of the dissolved substance is by no means determined solely by the total volume of the solution, as was the case in glycol. Indeed, the effect of temperature at constant volume is very pronounced and always results in an increase in apparent volume with rise of temperature. In FIGURE 18, we illustrate the effect of temperature on the

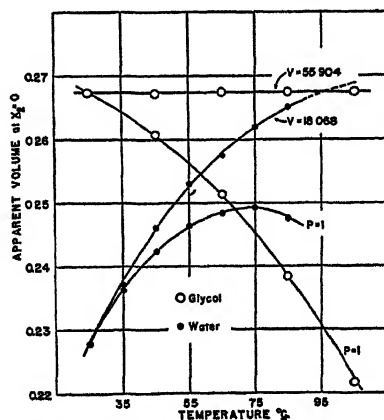


FIGURE 18. The effect of temperature on the apparent volume of sodium bromide in water and in glycol at constant pressure and constant volume.

apparent volume of sodium bromide in infinite dilution both in water and in glycol. At constant pressure, the apparent volume in glycol falls steadily with the rise of temperature; in water, it first rises anomalously and then

passes through the well-known maximum. At constant volume, the apparent volume of sodium bromide in glycol is essentially constant, while that in water rises continuously and appears to approach a value slightly in excess of the apparent volume in glycol. It is interesting to note that at 85° or 90° the compressibility of glycol is greater than that of water, whereas the reverse is the case at lower temperatures.

Conclusion. The effects of pressure, temperature, and concentration on the various thermodynamic properties of solutions discussed in this paper are all in qualitative agreement with the current theories of the molecular distribution in water. A comparison of the thermodynamic properties of sodium bromide in water and in glycol indicates that the contractions on mixing, the thermal expansions, and the energy-volume coefficients of the aqueous solutions at the lower temperatures are largely determined by the effects of short-range forces, and that the effects of ion-molecule interactions are only secondary. At higher temperatures the effects of short-range forces have a decreasing influence on these volumetric properties. Indeed, there is good reason to believe that, above 150°, the ion-molecule interactions and other long-range forces will play the significant part in determining the volumetric behavior of aqueous solutions and that simple regularities will be revealed that are masked at temperatures where most of our information is now available.

Bibliography

1. HALL, R. E. 1924. J. Wash. Acad. Sci. 14: 167.
2. GIBSON, R. E., & O. H. LOEFFLER. 1939. J. Am. Chem. Soc. 61: 2515.
3. GIBSON, R. E. 1937. J. Am. Chem. Soc. 59: 1521.
4. GIBSON, R. E., & O. H. LOEFFLER. 1941. J. Am. Chem. Soc. 63: 443.
5. RICHARDSON, J. M., A. B. ARONS, & R. R. HALVERSON. 1947. J. Chem. Phys. 15: 785.
6. GIBSON, R. E., & O. H. LOEFFLER. 1939. J. Phys. Chem. 43: 207.
7. GIBSON, R. E., & O. H. LOEFFLER. 1941. J. Am. Chem. Soc. 63: 2287.
8. GIBSON, R. E., & O. H. LOEFFLER. 1939. (See Equation 4) J. Am. Chem. Soc. 61: 2515.
9. GIBSON, R. E., & O. H. LOEFFLER. 1941. J. Am. Chem. Soc. 63: 443.
10. GIBSON, R. E., & O. H. LOEFFLER. 1941. J. Am. Chem. Soc. 63: 2287.
11. GIBSON, R. E., & O. H. LOEFFLER. 1941. J. Am. Chem. Soc. 63: 898.
12. GIBSON, R. E., & O. H. LOEFFLER. 1939. J. Am. Chem. Soc. 61: 2515.
13. GIBSON, R. E., & O. H. LOEFFLER. 1939. J. Am. Chem. Soc. 61: 2877.
14. GIBSON, R. E. 1934. J. Am. Chem. Soc. 56: 4.
15. KRITSCHESKY, I. R. 1938. Acta Physicochimica U.S.S.R.: 181.

EXTRAPOLATION OF APPARENT MOLAL PROPERTIES OF STRONG ELECTROLYTES

BY BENTON B. OWEN AND STUART R. BRINKLEY, JR.*

Sterling Chemistry Laboratory, Yale University, New Haven, Connecticut

Very soon after the appearance of the Debye-Hückel^{1, 2, 3} theory, it was recognized that the equations†

$$\log f_{\pm} = \frac{-\mathcal{S}_f c^{\frac{1}{2}}}{1 + \kappa a} + Bc \quad (1)$$

$$\mathcal{S}_f = 0.4343 \frac{1}{\nu} (\sum \nu_j z_j^2)^{\frac{1}{2}} \left(\frac{\pi N \epsilon^6}{1000 (kDT)^3} \right)^{\frac{1}{2}} \quad (2)$$

$$\kappa a = (\sum \nu_j z_j^2)^{\frac{1}{2}} \left(\frac{4\pi N \epsilon^2}{1000 kDT} \right)^{\frac{1}{2}} a c^{\frac{1}{2}} = A' c^{\frac{1}{2}} \quad (3)$$

adequately represented the available activity coefficient data⁵ for binary aqueous solutions of simple strong electrolytes, and that the inclusion of higher terms^{6, 7} in the development of the theory would permit its application to solvents of lower dielectric constants, and to electrolytes of more complex valence types. EQUATION 1, or its counterpart in terms of the osmotic coefficient, has long been generally employed for both interpolation and extrapolation of activity coefficient, or osmotic coefficient, data.

Equations for other thermodynamic properties can be derived from this equation by appropriate differentiation. Two such equations, which we will have occasion to use, represent the relative partial molal volume and relative partial molal heat content. These equations⁸

$$\bar{V}_2 - \bar{V}_2^0 = \frac{\mathcal{S}_v c^{\frac{1}{2}}}{1 - \kappa a} + \frac{W_v c}{(1 - \kappa a)^2} + K_v c \quad (4)$$

$$\bar{H}_2 - \bar{H}_2^0 = \frac{\mathcal{S}_H c^{\frac{1}{2}}}{1 + \kappa a} + \frac{W_H c}{(1 + \kappa a)^2} + K_H c \quad (5)$$

contain the theoretical limiting slopes

$$\mathcal{S}_v = 2.303 \nu RT \mathcal{S}_f \frac{3}{2} \left[\frac{\partial \ln D}{\partial P} - \frac{1}{3} \beta \right] \quad (6)$$

$$\mathcal{S}_H = -2.303 \nu RT^2 \mathcal{S}_f \frac{3}{2} \left[\frac{\partial \ln D}{\partial T} + \frac{1}{T} + \frac{1}{3} \alpha \right] \quad (7)$$

derived by differentiation of $\mathcal{S}_f c^{\frac{1}{2}}$. The symbols α and β represent the expansibility and compressibility, $-\partial \ln c / \partial T$ and $\partial \ln c / \partial P$, respectively. The terms containing the coefficients

* Present address: U. S. Bureau of Mines, Pittsburgh, Pa.

† For glossary of symbols not defined in this paper, consult HARNED, H. S., & B. B. OWEN.⁴

$$W_v = -2.303\nu RT \mathcal{S}_f A' \frac{1}{2} \left[\frac{\partial \ln D}{\partial P} - \beta - 2 \frac{\partial \ln a}{\partial P} \right] \quad (8)$$

$$W_H = 2.303\nu RT^2 \mathcal{S}_f A' \frac{1}{2} \left[\frac{\partial \ln D}{\partial T} + \frac{1}{T} + \alpha - 2 \frac{\partial \ln a}{\partial T} \right] \quad (9)$$

result from the factor $(1 + \kappa a)$ in EQUATION 1, and the last terms, K_{vc} and K_{Hc} , contain the derivatives of Bc .

EQUATION 5 has been employed, with an additional empirical term, by Harned and Ehlers,⁹ to represent relative partial molal heat contents derived from electromotive force data, but we are not aware of any application of EQUATION 4 except in very abbreviated forms. Later, we shall illustrate the evaluation of the terms in both of these complete EQUATIONS 4 and 5 after we have derived expressions for the corresponding apparent molal properties which are more closely related to measurable volume and heat effects.

In the derivation of these expressions, which, in alternative forms, will appear as EQUATIONS 22 and 23, and 35 and 36, detailed consideration will be given only to those terms corresponding to the first two terms of the right hand members of EQUATIONS 4 and 5. These theoretical terms contain no parameters which must necessarily be evaluated from volume or heat data on solutions, for the a -parameter is calculable by EQUATION 1 from activity coefficients. The final terms, $\frac{1}{2}K_{vc}$ and $\frac{1}{2}K_{Hc}$, contain empirically adjustable parameters which can, for the present, be evaluated only from volume and thermal effects of some kind. These terms correspond to K_{vc} and K_{Hc} in EQUATIONS 4 and 5 and are, in fact, derivable from them. They will be introduced in the last step of the derivations of our final equations. This procedure can be justified *a posteriori*, and it will simplify the treatment considerably.

Derivation of the Equations

For a solution containing n_1 moles of solvent and n_2 moles of electrolyte, it can be shown¹ that the Coulomb forces between the ions contribute the amount

$$\Delta F(eI) = -n_2 \frac{N\epsilon^3}{3D} \tau \kappa \sum_j z_j^2 \quad (10)$$

to the free energy of the solution. The function τ is defined by

$$\tau = \frac{3}{\kappa^3 a^3} \left[\frac{1}{2} \kappa^2 a^2 - \kappa a + \ln(1 + \kappa a) \right] \quad (11)$$

$$= 1 - \frac{3}{2} \kappa a + \frac{3}{2} \kappa^2 a^2 - \dots; \quad \kappa a \leq 1. \quad (11a)$$

The electrolytic contribution to the volume and enthalpy of the solution can be obtained from $\Delta F(eI)$ by the thermodynamic equations

$$\Delta V(el) = \left(\frac{\partial \Delta F(el)}{\partial P} \right)_T \quad (12)$$

$$\Delta H(el) = -T^2 \left(\frac{\partial \Delta F(el)/T}{\partial T} \right)_P \quad (13)$$

For the binary solution postulated, we can identify $\Delta V(el)$ and $\Delta H(el)$ with the total departures of the *ionized* solute from ideal behavior with respect to volume and enthalpy, and write

$$\Delta V(el) = (n_1 V_1 + n_2 \phi_v) - (n_1 V_1 + n_2 \phi_v^0) = n_2 (\phi_v - \phi_v^0) \quad (14)$$

$$\Delta H(el) = (n_1 H_1 + n_2 \phi_H) - (n_1 H_1 + n_2 \phi_H^0) = n_2 (\phi_H - \phi_H^0) \quad (15)$$

The combination of EQUATIONS 10 through 15 then yield the expressions

$$\phi_v - \phi_v^0 = -\frac{\epsilon^2 N}{3D} \sum \nu_j z_j^2 \left[\sigma \frac{\partial \ln \kappa a}{\partial P} - \tau \frac{\partial \ln Da}{\partial P} \right] \kappa \quad (16)$$

$$\phi_H - \phi_H^0 = T \frac{\epsilon^2 N}{3D} \sum \nu_j z_j^2 \left[\sigma \frac{\partial \ln \kappa a}{\partial T} - \frac{\tau}{T} - \tau \frac{\partial \ln Da}{\partial T} \right] \kappa \quad (17)$$

where the function σ is defined by

$$\sigma = \frac{\partial(\tau \kappa a)}{\partial \kappa a} \quad (18)$$

It follows from EQUATIONS 18 and 11 that

$$\sigma = \frac{3}{\kappa^2 a^3} \left[1 + \kappa a - \frac{1}{1 + \kappa a} - 2 \ln(1 + \kappa a) \right] \quad (19)$$

$$= 1 - \frac{3}{2} \kappa a + \frac{3}{8} \kappa^2 a^2 - \dots; \quad \kappa a \leq 1 \quad (19a)$$

and EQUATION 3 leads to

$$\frac{\partial \ln \kappa a}{\partial P} = -\frac{1}{2} \frac{\partial \ln D}{\partial P} + \frac{1}{2} \beta + \frac{\partial \ln a}{\partial P} \quad (20)$$

$$\frac{\partial \ln \kappa a}{\partial T} = -\frac{1}{2} \frac{\partial \ln D}{\partial T} - \frac{1}{T} - \frac{1}{2} \alpha + \frac{\partial \ln a}{\partial T} \quad (21)$$

By combining EQUATIONS 16 through 21, properly rearranging terms in τ and σ , and introducing the empirically adjustable terms, $\frac{1}{2}K_v c$ and $\frac{1}{2}K_H c$, referred to above, we obtain

$$\phi_v - \phi_v^0 = \frac{2}{3} \mathfrak{S}_v \tau c^{\frac{1}{2}} + \frac{1}{2} W_v \theta c + \frac{1}{2} K_v c \quad (22)$$

$$\phi_H - \phi_H^0 = \frac{2}{3} \mathfrak{S}_H \tau c^{\frac{1}{2}} + \frac{1}{2} W_H \theta c + \frac{1}{2} K_H c \quad (23)$$

The function θ is defined by

$$\theta = \frac{4}{3} \left(\frac{\tau - \sigma}{\kappa a} \right) \quad (24)$$

$$= 1 - \frac{3}{2} \kappa a + \frac{1}{8} \kappa^2 a^2 - \dots; \quad \kappa a \leq 1. \quad (24a)$$

It is useful to note the similarity and differences between EQUATIONS 4 and 5 for partial molal properties and EQUATIONS 22 and 23 for the corresponding apparent molal properties. The effects of the α -parameter appear as the functions τ and θ in the latter equations and as $(1 + \alpha a)^{-1}$ and $(1 + \kappa a)^{-2}$ in the former. The characteristic coefficients of the various terms are, however, identical except for the numerical factors $\frac{2}{3}$ and $\frac{1}{2}$.

To justify the introduction of the terms $\frac{1}{2}K_v c$ and $\frac{1}{2}K_H c$ into EQUATIONS 22 and 23, it is sufficient to show that these equations can be derived from EQUATIONS 4 and 5. It follows from definitions that

$$\phi_v - \phi_v^0 = \frac{1}{m} \int_0^m (\bar{V}_2 - \bar{V}_2^0) dm \quad (25)$$

$$\phi_H - \phi_H^0 = \frac{1}{m} \int_0^m (\bar{H}_2 - \bar{H}_2^0) dm \quad (26)$$

$$m = \frac{c}{d_0} \left[1 - \frac{c\phi_v}{1000} \right]^{-1}. \quad (27)$$

For concentrations so low that $c\phi_v/1000$ is negligible compared with unity, it is not difficult to substitute EQUATIONS 4 and 27 into EQUATION 25, or EQUATIONS 5 and 27 into 26, and obtain EQUATIONS 22 or 23, by integration. For higher concentrations, the forms of the general integrals are too complicated to be of much practical use, but it can be shown by expansion in series that the terms $\frac{1}{2}K_v c$ and $\frac{1}{2}K_H c$ satisfactorily represent the integrals,

$$\frac{1}{m} \int_0^m K_v c dm \quad \text{and} \quad \frac{1}{m} \int_0^m K_H c dm$$

for simple electrolytes at concentrations up to about 1 molar.

EQUATIONS 22 and 23 can be employed for extrapolation by rewriting them in the forms

$$[\phi_v - \frac{2}{3}\mathfrak{S}_v\tau c^{\frac{1}{2}} - \frac{1}{2}W_v\theta c] = \phi_v^0 + \frac{1}{2}K_v c \quad (28)$$

$$[\phi_H - \phi'_H - \frac{2}{3}\mathfrak{S}_H\tau c^{\frac{1}{2}} - \frac{1}{2}W_H\theta c] = \phi_H^0 - \phi'_H + \frac{1}{2}K_H c. \quad (29)$$

In the last equation, ϕ'_H represents the apparent molal heat content at some convenient experimental concentration to which all heats are temporarily referred pending extrapolation. It is clear that plots of the left hand members of these equations should be straight lines. The intercepts of these lines at zero concentration are ϕ_v^0 and $\phi_H^0 - \phi'_H$, and the slopes are $\frac{1}{2}K_v$ and $\frac{1}{2}K_H$. This procedure has the advantage that it evaluates all of the coefficients, both theoretical and adjustable, which appear in EQUATIONS 22 and 23 for the apparent molal properties and in EQUATIONS 4 and 5 for the partial molal properties. It has the disadvantage of being unnecessarily laborious when one is interested primarily in extrapolation.

More convenient expressions for extrapolation can be derived from EQUATIONS 16 and 17 by making use of the relation

$$2\tau + \sigma = \frac{3}{1 + \kappa a} \quad (30)$$

in regrouping the terms within the brackets. Thus, by combining EQUATION 30 with EQUATIONS 16 through 21, and introducing the empirically adjustable terms, $\frac{1}{2}K_v c$ and $\frac{1}{2}K_H c$, we obtain

$$\begin{aligned} \phi_v - \phi_v^0 &= 2.303\nu RT \bar{S}_f \left[\frac{1}{1 + \kappa a} \frac{\partial \ln D}{\partial P} - \frac{\sigma}{3} \beta \right] c^{\frac{1}{2}} \\ &+ 2.303\nu RT \bar{S}_{fA'} \left[\frac{\partial \ln a}{\partial P} \right] \theta c + \frac{1}{2} K_v c \end{aligned} \quad (31)$$

$$\begin{aligned} \phi_H - \phi_H^0 &= -2.303\nu RT^2 \bar{S}_f \left[\frac{1}{1 + \kappa a} \left\{ \frac{\partial \ln D}{\partial T} + \frac{1}{T} \right\} + \frac{\sigma}{3} \alpha \right] c \\ &+ 2.303\nu RT^2 \bar{S}_{fA'} \left[\frac{\partial \ln a}{\partial T} \right] \theta c + \frac{1}{2} K_H c. \end{aligned} \quad (32)$$

The practical advantages of these equations over EQUATIONS 22 and 23 will become apparent when we undertake to evaluate the various terms within the brackets. In the first place, we will assume* that both $\partial \ln a / \partial P$ and $\partial \ln a / \partial T$ are zero. This completely eliminates the terms in θc from EQUATIONS 31 and 32, but merely alters the coefficients, W_v and W_H , of the terms in θc of EQUATIONS 22 and 23. In the second place, the coefficients of $c^{\frac{1}{2}}$ in EQUATIONS 31 and 32 must reduce to $\frac{2}{3}\bar{S}_v$ and $\frac{2}{3}\bar{S}_H$ at infinite dilution, so it will be convenient to introduce these limiting slopes by means of EQUATIONS 6 and 7 and obtain

$$\phi_v - \phi_v^0 = \frac{2}{3} \bar{S}_v \Omega_v c^{\frac{1}{2}} + \frac{1}{2} K_v c \quad (33)$$

$$\phi_H - \phi_H^0 = \frac{2}{3} \bar{S}_H \Omega_H c^{\frac{1}{2}} + \frac{1}{2} K_H c \quad (34)$$

Consequently, our final equations in the forms which we find most convenient for extrapolation are

$$[\phi_v - \frac{2}{3} \bar{S}_v \Omega_v c^{\frac{1}{2}}] = \phi_v^0 + \frac{1}{2} K_v c \quad (35)$$

$$[\phi_H - \frac{2}{3} \bar{S}_H \Omega_H c^{\frac{1}{2}}] = \phi_H^0 - \phi_H' + \frac{1}{2} K_H c. \quad (36)$$

As before, ϕ_H' is the apparent molal heat content at some temporary reference concentration. Plots of the left hand members of these equations against c should be straight lines and permit easy extrapolation. It will appear later that the labor of calculating $\bar{S}_v \Omega_v$ and $\bar{S}_H \Omega_H$ is less than that involved in evaluating the corresponding pairs of terms in EQUATIONS 28 and 29.

The new functions, Ω_v and Ω_H , are defined by

$$\Omega_v = \left[\frac{1}{1 + \kappa a} \frac{\partial \ln D}{\partial P} - \frac{\sigma}{3} \beta \right] \left[\frac{\partial \ln D}{\partial P} - \frac{\sigma}{3} \beta \right]^{-1} \quad (37)$$

* The α -parameter is known to be very insensitive to temperature (See ref. 9, for example), and its variation with pressure is unknown.

$$\Omega_H = \left[\frac{1}{1 + \kappa a} \left\{ \frac{\partial \ln D}{\partial T} + \frac{1}{T} \right\} + \frac{\sigma}{3} \alpha \right] \left[\frac{\partial \ln D}{\partial T} + \frac{1}{T} + \frac{1}{3} \alpha \right]^{-1}. \quad (38)$$

Although Ω_v and Ω_H become unity at infinite dilution, they decrease rather rapidly with increasing concentration, and this decrease is the more rapid the more dilute the solution. For example, in 0.01 normal KCl solution, Ω_v is about 0.8. This is a result of practical importance, for it shows that, at almost the lower limit of useful measurements of ϕ_v , the *theoretical* value of $\partial\phi_v/\partial c^{\frac{1}{2}}$ is already 20 per cent lower than the limiting value, $\frac{2}{3}\mathfrak{S}_v$. The commonly observed linearity of plots of ϕ_v against $c^{\frac{1}{2}}$ has proved so convenient and satisfactory for the empirical treatment^{10, 11} of density data that it seems to have removed the incentive for a proper application of the Debye-Hückel theory. When very precise data became available in dilute solutions, it was found¹² that the addition of a term linear in c was necessary, but the α -parameter was disregarded. In all of these previous studies of ϕ_v as a function of concentration, it was found that the observed value of $\partial\phi_v/\partial c^{\frac{1}{2}}$ differed considerably from $\frac{2}{3}\mathfrak{S}_v$ calculated by EQUATION 6. We will show later that this apparent disagreement with theory was due to the incomplete application of theory brought about by disregarding the α -parameter. Scatchard and Epstein¹³ appear to have been the only writers who attempted to use the α -parameter in representing ϕ_v , but the development of their final equation involved so many approximations that they were led to conclude that "... the form we have chosen for the electrostatic term is not helpful, for the deviation from the limiting law is smaller than that from our expression, and it may be represented more closely by a term linear in m ."

The treatment of accurate heat data has also been incomplete, but in quite a different manner. In this case, the α -parameter was relied upon wholly to represent the departures from the limiting law, and the linear term in c was disregarded. Here, also, the experimental data appeared to be represented^{14, 15} satisfactorily. Unfortunately, the magnitudes of the α -parameters required were in disagreement with the magnitudes derived¹⁶ from activity coefficient data for the same salts by EQUATION 1. Here, again, this apparent disagreement with theory is due to its incomplete application, which involves, in this case, neglect of the term linear in c .

By means of a few simple applications, we propose to show that both the α -parameter and the term $\frac{1}{2}Kc$ make important contributions to the departures from the limiting laws within the experimental range of "dilute" solutions, and that the neglect of neither one nor the other is permissible in a treatment which pretends to be reasonably complete. In order to do this, however, we must carefully consider the numerical evaluation of the theoretical coefficients \mathfrak{S}_v , \mathfrak{S}_H , W_v , Ω_H , etc.

Evaluation of the Coefficients

The coefficients \mathfrak{S}_H and W_H can be calculated with reasonable accuracy because the temperature variation of the dielectric constant and the expansi-

bility of water have been the subject of several careful studies. In calculating the values recorded in TABLE 1, we used the equation of Wyman and

TABLE 1
THEORETICAL COEFFICIENTS FOR UNI-UNIVALENT ELECTROLYTES*

t°	\bar{S}_f	$A' \frac{10^{-8}}{a}$	$\frac{2}{3}\bar{S}_v$	$-\frac{1}{2}W_v \frac{10^{-8}}{a}$	$\frac{2}{3}\bar{S}_H$	$-\frac{1}{2}W_H \frac{10^{-8}}{a}$	h
0	.4883	.3241	2.467	.0594	339.5	28.73	-.0222
5	.4921	.3249	2.456	.0593	362.3	29.13	+.0051
10	.4960	.3258	2.457	.0595	387.0	29.78	.0275
15	.5002	.3267	2.468	.0599	413.5	30.65	.0462
20	.5046	.3276	2.488	.0606	441.7	31.70	.0619
25	.5091	.3286	2.517	.0615	471.8	32.93	.0752
30	.5139	.3297	2.555	.0626	503.7	34.30	.0866
35	.5189	.3307	2.601	.0640	537.5	35.85	.0966
40	.5241	.3318	2.655	.0655	573.3	37.54	.1054
45	.5295	.3330	2.717	.0673	611.4	39.38	.1130
50	.5351	.3341	2.787	.0692	652.1	41.43	.1196
12.5	.4981	.3262	2.461	.0597	400.0	30.19	.0373
18	.5028	.3273	2.479	.0603	430.2	31.27	.0559
Factor	$\frac{2}{v}f^{3/2}$	$f^{1/2}$	$\frac{2}{v}f^{3/2}$	f^2	$\frac{2}{v}f^{3/2}$	f^2	1

* For electrolytes of other valence types, multiply the figures in the table by the proper factor given in the last line. f represents the valence factor $\frac{1}{2}\Sigma v_i z_i^2$.
In using this table, it is necessary to express ϕ_v in cc. per mole, ϕ_H in 15° calories per mole, and a in cm.

Ingalls¹⁷ to evaluate $\partial \ln D/\partial T$, and the Tilton and Taylor¹⁸ equation to evaluate α from the data of Chappuis.¹⁹ For the fundamental physical constants, ϵ , N , k , T , etc., we used the values recommended by Birge.²⁰

For evaluating the coefficients \bar{S}_v and W_v , several series of measurements of the compressibility of water are available. We have chosen the results of Gibson,²¹ which cover a wide temperature range and have been expressed in terms of the Tait²² equations

$$\frac{v}{v^1} = 1 - (C/v^1) \log \left(\frac{B + P}{B + 1} \right) \quad (39)$$

$$\beta = \frac{0.4343(C/v^1)}{B + 1} \quad (39a)$$

where v and v^1 represent specific volumes at the pressure P and at 1 bar, respectively. The quantity $0.4343(C/v^1)$ is equal to 0.1368 for water, and can be considered independent of temperature. B , expressed in bars, varies with temperature according to the equation²³

$$B = 2996 + 7.5554(t - 25) - 0.17814(t - 25)^2 + 0.000608(t - 25)^3. \quad (40)$$

A few years ago, the authors demonstrated²⁴ that similar equations

$$\frac{D^1}{D} = 1 - (AD^1) \log \left(\frac{B + P}{B + 1} \right) \quad (41)$$

$$\frac{\partial \ln D}{\partial P} = \frac{0.4343(AD^1)}{B + 1} \quad (41a)$$

could express the pressure variation of the dielectric constants of a wide variety of liquids. D and D^1 represent the dielectric constant at P and at 1 bar, respectively. It was also shown that the parameter (AD^1) , which is analogous to (C/v^1) of EQUATION 39, is essentially independent of temperature, and that the value of B for any given liquid is the same as that derived from the Tait Equation (39a) and compressibility data.

Unfortunately, the only extensive measurements of the dielectric constant of water under pressure are those of Kyropoulos²⁵ at the single temperature 20°, and at only six pressures above atmospheric. Until these measurements are repeated and extended to other temperatures, we seem to have no choice but to employ EQUATION 41 and assume that water behaves enough like other liquids for (AD^1) to be independent of temperature. Accordingly, using the value of B given by EQUATION 40 at 20°, converting the pressures to bars, and multiplying the dielectric constants by 80.362/80.790 to correct to modern standards,¹⁷ the data of Kyropoulos lead to the value 0.1754 for the quantity $0.4343(AD^1)$. This value was used in the evaluation of \bar{S}_\circ and \bar{W}_\circ at all temperatures recorded in TABLE 1.

An interesting and very convenient consequence of the common denominator of EQUATIONS 39a and 41a is that the function Ω_\circ is given by

$$\Omega_\circ = \frac{1.3513}{1 + \kappa a} - 0.3513\sigma \quad (42)$$

at all temperatures. The corresponding function, Ω_H , which can be written in the form

$$\Omega_H = \frac{1 + h}{1 + \kappa a} - h\sigma \quad (43)$$

varies with the temperature. Values of h are included in TABLE 1.

Small tables of the functions σ and τ are available in the literature,^{1, 4, 15, 26} but the intervals are too large for accurate interpolation. Scatchard and Epstein¹⁸ have constructed a table of $Z = \sigma\kappa a/3$ at even intervals of $y = \kappa a/(1 + \kappa a)$ which is suitable for linear interpolation, and which permits the calculation of both σ and τ . One of us (Brinkley) has prepared an extensive table of σ and τ at even intervals of κa which may be published elsewhere.

Illustrative Examples of Extrapolation

FIGURE 1 illustrates the extrapolation of the heats of dilution of sodium chloride, a typical strong electrolyte. The circles represent apparent molal heat contents, relative to ϕ'_H at $c = 0.1$, calculated from the data of Robinson.²⁷ The curve drawn through the circles is the customary plot against $c_i^{\frac{1}{2}}$ drawn so as to merge with the limiting law (L. L.) just below the experi-

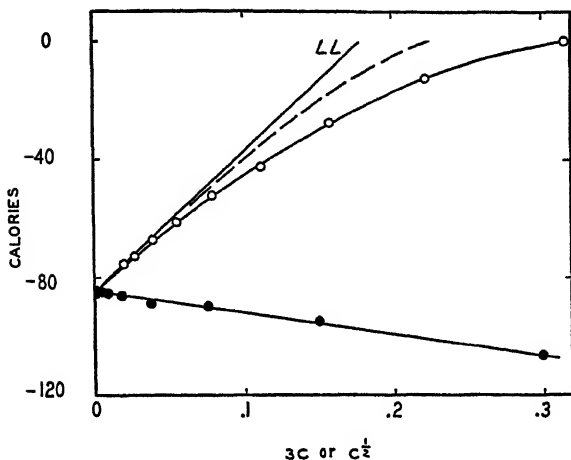


FIGURE 1. Relative apparent molal heat content of sodium chloride at 25°. —O— $\phi_H - \phi_H'$ against $c^{1/2}$; — $\phi_H - \phi_H'$ against $\Omega_H c^{1/2}$; —●— $\phi_H - \phi_H' - 471.8\Omega_H c^{1/2}$ against c . ϕ_H' at $c = 0.1$.

mental concentration range. The broken line indicates where this curve would lie if $\Omega_H c^{1/2}$ instead of $c^{1/2}$ had been used as abscissa. In the calculation of Ω_H , and of Ω_ν in all subsequent figures, we used values of the α -parameters which Robinson and Harned¹⁶ obtained by fitting EQUATION 1 to the best available activity coefficient data.

The filled-in circles in FIGURE 1 represent the left hand member of EQUATION 36 plotted against $3c$.^{*} It is clear that a straight line represents these points over the entire experimental range and permits a very satisfactory extrapolation. From the slope and intercept of this line, we find that

$$\phi_H - \phi_H^0 = 471.8\Omega_H c^{1/2} + 217c; \quad a = 4.0 \times 10^{-8} \quad (44)$$

for sodium chloride at 25°.

FIGURE 2 illustrates the extrapolation of the apparent molal volume of sodium chloride. The circles represent values of ϕ_ν calculated from the data of Kruis²⁸ and plotted against $c^{1/2}$. The curve through these points is drawn to merge with the limiting law. The S-shaped nature of this curve is barely discernable† on the plot. It has been common practice to overlook this inflection, and to draw the best straight line through a very considerable concentration range, and to use this line for extrapolation. As can be seen from the figure, this practice does not introduce a serious error into the evaluation of ϕ_ν^0 for the examples chosen, but it completely obscures an important feature of the concentration dependence of ϕ_ν . The broken line shows where the plot of ϕ_ν against $\Omega_\nu c^{1/2}$ would lie. This curve does not pass through an inflection, and merges more rapidly with the limiting law than

^{*} The scale factor 3 is introduced for convenience in this particular plot.

† This feature is more clearly observed in the plot of ϕ_ν against $c^{1/2}$ for strontium chloride. See H. S. HARNED & B. B. OWEN *Op. cit.*: 256.

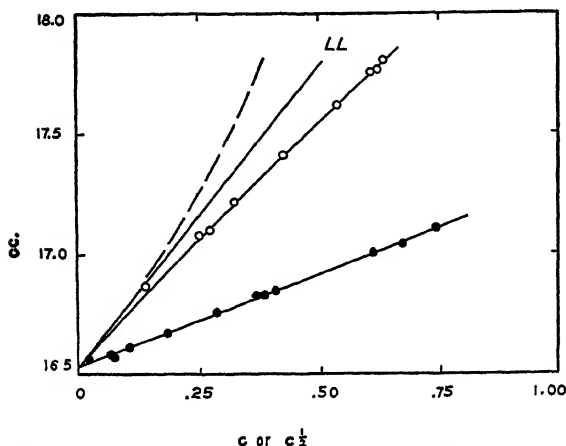


FIGURE 2. Apparent molal volume of sodium chloride at 25°. —○— ϕ_v against $c^{\frac{1}{2}}$; — — ϕ_v against $\Omega_p c^{\frac{1}{2}}$; —●— $\phi_v - 2.517\Omega_p c^{\frac{1}{2}}$ against c .

the plot of ϕ_v against $c^{\frac{1}{2}}$. The same value¹⁶ of the a -parameter, 4.0×10^{-8} , was used as in FIGURE 1.

The inked-in circles in FIGURE 2 represent the left hand member of EQUATION 35 plotted against c . The straight line through these points leads to the equation

$$\phi_v = 16.538 + 2.517\Omega_p c^{\frac{1}{2}} + 0.76c; \quad a = 4.0 \times 10^{-8} \quad (45)$$

for sodium chloride at 25°.

A similar set of curves for potassium chloride is exhibited in FIGURE 3,

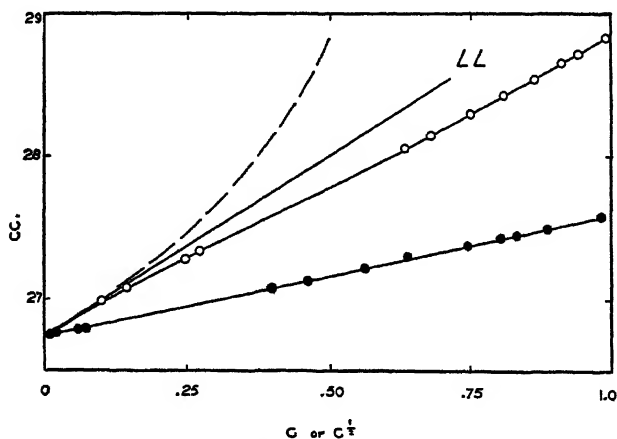


FIGURE 3. Apparent molal volume of potassium chloride at 25°. —○— ϕ_v against $c^{\frac{1}{2}}$; — — ϕ_v against $\Omega_p c^{\frac{1}{2}}$; —●— $\phi_v - 2.517\Omega_p c^{\frac{1}{2}}$ against c .

based upon the very precise data of Geffcken and Price.¹² The concentration range in this case is greater than that of FIGURE 2, and the plot of $\phi_v - 2.517 \Omega_v c^{\frac{1}{2}}$ against c is a straight line over the entire range. The equation of this line is

$$\phi_v = 26.742 + 2.517 \Omega_v c^{\frac{1}{2}} + 0.85c; \quad a = 3.8 \times 10^{-8} \quad (46)$$

for potassium chloride at 25°. The independently determined¹⁶ value of the a -parameter, 3.8×10^{-8} , is remarkably satisfactory in representing the data.

It should be pointed out, however, that, from a purely numerical point of view, the data for sodium chloride, potassium chloride, hydrochloric acid, and several other uni-univalent electrolytes can be treated almost equally well by neglecting the a -parameter and using an empirical slope of about 1.86 to 1.90, or by retaining the a -parameter and using the theoretical limiting slope, 2.517. The intercepts, ϕ_v^0 , are not the same by the two methods of extrapolation. For sodium chloride, potassium chloride, and hydrochloric acid, the empirical extrapolations yield values of ϕ_v^0 about 0.07 cc. higher than the theoretical. For complex valence type electrolytes, this difference would probably be much greater. We have omitted such electrolytes from our illustrative extrapolations so as to avoid complications due to the extended terms.^{6, 7, 15}

Finally, FIGURE 4 shows, on a magnified scale, the extrapolation of ϕ_v

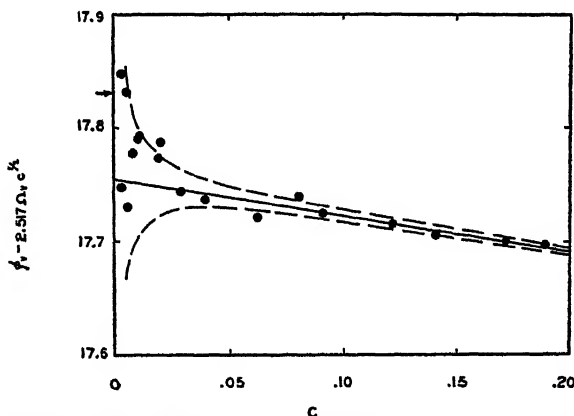


FIGURE 4. Apparent molal volume of hydrochloric acid at 25°. — — Uncertainty corresponding to an experimental error of $\pm 5 \times 10^{-7}$ in the density.

for hydrochloric acid calculated from the data of Redlich and Bigeleisen.²⁰ With the a -parameter taken as 4.4×10^{-8} from activity coefficient data,¹⁶ the straight line extrapolation in the figure leads to the equation

$$\phi_v = 17.755 + 2.517 \Omega_v c^{\frac{1}{2}} - 0.32c; \quad a = 4.4 \times 10^{-8} \quad (47)$$

for hydrochloric acid at 25°. The broken lines are error curves showing the deviations from the straight line which would result from an error of $\pm 5 \times 10^{-7}$ in the original density measurements. Redlich and Bigeleisen chose to represent their results by the equation

$$\phi_v = 17.830 + 1.86 c^{\frac{1}{2}} - 1.15 c; \quad a = 0 \quad (48)$$

which neglects the a -parameter and makes use of an empirical limiting slope, 1.86, based upon their data. Their extrapolation leads to the value of ϕ_v^0 indicated by the arrow on FIGURE 4. It relies largely upon the points at the highest dilutions grouped on the high side of the error curve, and requires a much lower value of the limiting slope³⁰ than the theoretical value, 2.517, calculated from the properties of water.

This leads us to conclude that, for precise extrapolation of apparent molal properties, one should not depend wholly upon the limiting law and results at very high dilution, but should apply a more complete theoretical treatment of the data at moderate dilutions.

Bibliography

1. DEYBE, P., & E. HÜCKEL. 1923. Phys. Z. 24: 185.
2. DEYBE, P. 1924. Ibid. 25: 97.
3. HÜCKEL, E. 1925. Ibid. 26: 93.
4. HARNED, H. S., & B. B. OWEN. 1943. *The Physical Chemistry of Electrolytic Solutions*. Reinhold. New York.
5. HARNED, H. S., & G. ÅKERLÖF. 1926. Ibid. 27: 411.
6. GRONWALL, T. H., V. K. LAMER, & K. SANDVED. 1929. Ibid. 29: 358.
7. LAMER, V. K., T. H. GRONWALL, & L. J. GREIFF. 1931. J. Phys. Chem. 35: 2245.
8. HARNED, H. S., & B. B. OWEN. *Op. Cit.*, sections 8 and 9 of Chapter III.
9. HARNED, H. S., & R. W. EHLERS. 1933. J. Am. Chem. Soc. 55: 2179.
10. MASSON, D. O. 1929. Phil. Mag. 8(7): 218.
11. SCOTT, A. F. 1931. J. Phys. Chem. 35: 2315.
12. GEFFCKEN, W., & D. PRICE. 1934. Z. phys. Chem. B 26: 81.
13. SCATCHARD, G., & L. F. EPSTEIN. 1942. Chem. Rev. 30: 211.
14. LANGE, E., & H. STREECK. 1931. Naturwissenschaften 19: 359.
15. FALKENHAGEN, H. 1934. *Electrolytes* (Section 53). (Trans. R. P. BELL.) Clarendon Press. Oxford.
16. ROBINSON, R. A., & H. S. HARNED. 1941. Chem. Rev. 28: 419.
17. WYMAN, J., JR., & E. N. INGALLS. 1938. J. Am. Chem. Soc. 60: 1182.
18. TILTON, L. W., & J. K. TAYLOR. 1937. J. Res. N. B. S. 18: 205.
19. CHAPPUIS, P. 1907. Trav. & Mém. Bur. Internat. Poids & Mesures 13: D40.
20. BIRGE, R. T. 1941. Rev. Mod. Phys. 13: 233.
21. GIBSON, R. E. 1934. J. Am. Chem. Soc. 56: 4. 1935. Ibid. 57: 284.
22. TAIT, P. G. 1888. *Physics and Chemistry of the Voyage of H. M. S. Challenger*. II (IV): Suppl. LXI.
23. OWEN, B. B. 1944. J. Chem. Ed. 21: 59.
24. OWEN, B. B., & S. R. BRINKLEY, JR. 1943. Phys. Rev. 64: 32.
25. KYROPOULOS, S. 1926. Z. Phys. 40: 507.
26. FOWLER, R. H., & E. A. GUGGENHEIM. 1939. *Statistical Thermodynamics*. 399. University Press. Cambridge.
27. ROBINSON, A. L. 1932. J. Am. Chem. Soc. 54: 1311.
28. KRUIS, A. 1936. Z. phys. Chem. B34: 1.
29. REDLICH, O., & J. BIGEISEN. 1942. J. Am. Chem. Soc. 64: 758.
30. REDLICH, O. 1940. J. Phys. Chem. 44: 619.

THE SURFACE TENSIONS OF AQUEOUS SULFURIC ACID SOLUTIONS

BY T. F. YOUNG AND S. R. GRINSTEAD

George Herbert Jones Chemistry Laboratory, University of Chicago, Chicago, Illinois

The surface tension-composition isotherms of aqueous sulfuric acid solutions contain several interesting anomalies, shown in FIGURE 1. The curve for 18°C. was determined long ago by Röntgen and Schneider,¹ who employed the capillary height method in an investigation of solutions of a large number of electrolytes. Their measurements showed that the surface tensions of dilute solutions differ very little from that of pure water at 18°C. and that, as concentration is increased, the surface tension rises rapidly to a broad maximum and then drops sharply. The curves for 0°, 30°, and 50°C., shown in FIGURE 1, were obtained from the work of Morgan and Davis,² who made three long series of drop weight determinations. Though their work was done before the drop weight method had been developed,³ they recorded their precise observations in sufficient detail to permit calculation of surface tensions (Young and Harkins⁴). The measurements of Morgan and Davis revealed an interesting minimum in the surface tension curve at 0°C. Their curves at higher temperatures, however, do not exhibit minima, and lead to the question: Does the minimum merely become less prominent as temperature rises, or does it disappear entirely? That the maximum becomes more prominent and moves to higher concentrations as temperature rises is shown by the four curves of FIGURE 1.

It is the purpose of this paper to describe an attempt to explain these anomalies and to present semiquantitative theoretical calculations of the positions and magnitudes of the minima and the maxima at various temperatures.

Electrolytic Dissociation as an Explanation of the Properties of Sulphuric Acid Solutions. Several properties of dilute and moderately concentrated sulfuric acid solutions have been successfully explained by treatment of those solutions as mixtures of solvent, sulfate ion, bisulfate ion, and hydrogen ions. Sherrill and Noyes⁵ have explained the conductance of dilute solutions of sulfuric acid by treating them in that way, and were able, thus, to make a satisfactory determination of the dissociation constant of the HSO_4^- ion at 25°C.

A similar point of view is to regard the solute, H_2SO_4 , as a mixture of the three solutes: (1) an acid which is completely dissociated into sulfate ions and hydrogen ions; (2) an acid completely dissociated into hydrosulfate ions and hydrogen ions; and (3) a solute which is not ionized at all. To facilitate a discussion in these terms, the formula $\text{H}\cdot\text{H}\cdot\text{SO}_4$ will be used to represent that portion of the sulfuric acid which is dissociated into hydrogen ions and

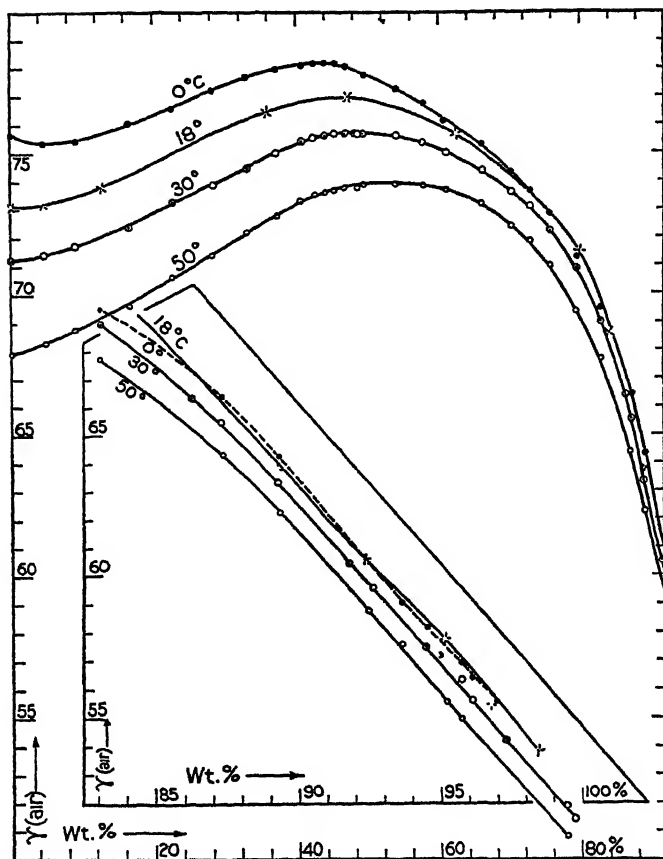


FIGURE 1. Surface tension (dyne/cm) of aqueous sulfuric acid solutions, versus weight per cent acid

sulfate ions; the symbol $\text{H} \cdot \text{HSO}_4$ will denote the other ionized solute; and the formula $\cdot \text{HSO}_4$ will denote the undissociated acid. The concentration of $\text{H} \cdot \text{H} \cdot \text{SO}_4$ is the SO_4^{2-} concentration, the concentration of $\text{H} \cdot \text{HSO}_4$ is the concentration of HSO_4^- . The customary formula, H_2SO_4 , will denote as usual the mixture of molecular and ionic species actually present in a real solution of H_2SO_4 in water. The stoichiometric concentration of H_2SO_4 is obviously the sum of the respective concentrations of $\text{H} \cdot \text{H} \cdot \text{SO}_4$, $\text{H} \cdot \text{HSO}_4$, and $\cdot \text{HSO}_4$.

Klotz and Eckert⁶ have been able in this manner to explain the rapid variation with dilution of the apparent molal volume of sulfuric acid in dilute solutions. They derived a curve to represent the sum of the apparent molal volumes of SO_4^{2-} and twice that of H^+ , i.e., a curve to represent the apparent molal volume of $\text{H} \cdot \text{H} \cdot \text{SO}_4$ (FIGURE 2). Another curve was determined which shows the variation with ionic strength of the apparent molal

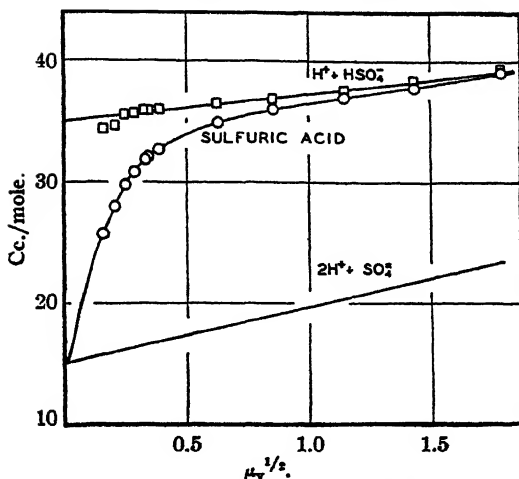


FIGURE 2. Apparent molal volumes (cm^3/mole) of sulfuric acid. The lowest line was calculated for $\text{H}\cdot\text{H}\cdot\text{SO}_4$. Experimental points, \bigcirc , for H_2SO_4 were derived from measured densities. The points \square were calculated for $\text{H}\cdot\text{HSO}_4$ by means of EQUATION 1 and available values of α_2 from the same densities. As expected, those points lie close to a straight line drawn (empirically) to represent apparent molal volumes of $\text{H}\cdot\text{HSO}_4$.

volume of $\text{H}\cdot\text{HSO}_4$. With these, they computed a third curve for the mixtures which exist in real solutions of H_2SO_4 . They employed the equation

$$\Phi(\text{H}_2\text{SO}_4) = \alpha_2\Phi(\text{H}\cdot\text{H}\cdot\text{SO}_4) + [1 - \alpha_2]\Phi(\text{H}\cdot\text{HSO}_4). \quad (1)$$

Here, Φ represents the respective apparent molal volumes of $\text{H}\cdot\text{H}\cdot\text{SO}_4$ and $\text{H}\cdot\text{HSO}_4$ and α_2 the degree of dissociation of the HSO_4^- ion. The top line of FIGURE 2 shows the agreement of these authors' beautiful interpretation with their precise experimental data. A similar but somewhat more complex calculation has been made of the heats of dilution of H_2SO_4 in solutions more dilute than 0.2 molal.⁷ Both of these calculations are based upon the observation that the apparent molal volume, Φ , and the apparent molal heat content, Φ_H , of each solute are functions of the total ionic strength. Each apparent volume or heat content of the solute is regarded simply as the sum of the individual apparent molal quantities, each multiplied by the respective molality.

The Surface Tensions of Ternary Solutions. If a similar treatment is to be possible for surface tension, it is necessary that a method be available for the calculation of the surface tensions of mixtures. Belton⁸ has recently investigated that problem and found that the difference between the surface tension, γ , of a solution and that of the pure solvent, γ_0 , is approximately equal to the sum of the increments produced by each of the respective solutes when present alone in the same solvent. Thus

$$\Delta\gamma \equiv \gamma - \gamma_0 \approx \sum \Delta\gamma_i \quad (2)$$

Here, $\Delta\gamma$, represents the surface tension increment due to the i^{th} constituent in a simple binary solution in which the concentration of that solute is the same as it is in the complex mixture. (Note the absence of an ionic strength effect.) When hydrogen ions and a second species of positive ion are both present, EQUATION 2 fails. Hydrochloric acid, for example, reduces the surface tension of a sodium chloride solution much more than it lowers the surface tension of pure water. EQUATION 2 will be used here only for solutions which contain but one positive ion.

Independent Contributions to Surface Tension of the Ionic Constituents of Binary Solutions. An additional principle is required for the complete treatment of real solutions of H_2SO_4 as mixtures of the solutes: $\text{H} \cdot \text{HSO}_4$ and $\text{H} \cdot \text{H} \cdot \text{SO}_4$. It is necessary to determine the surface tension of a hypothetical solution containing the solvent and a single one of the (hypothetical) solutes at specified concentrations. It is necessary, for example, to estimate the surface tension of solutions containing only the solvent and $\text{H} \cdot \text{H} \cdot \text{SO}_4$, i.e., containing only hydrogen ions and sulfate ions, not bisulfate ions. Such a principle (EQUATION 5) was evolved during an examination of representative surface tension data available in the *International Critical Tables*.⁴

FIGURE 3 exhibits a set of typical surface tension increment isotherms for 20°C. The lines characteristic of typical strong salts are nearly linear from very small to comparatively large molalities. The slopes of those lines are nearly independent of temperature; indeed, the temperature variations of the slopes are within the precisions of the respective measurements. The isotherms for HCl and HNO_3 exhibit two conspicuous peculiarities. The slopes of those two curves are negative within the range investigated and change somewhat with composition. The variations of the surface tensions of acid solutions in the concentration range to which the theory of Onsager and Samaras⁹ is applicable have not been determined. Fortunately, data for such high dilutions are not necessary for our immediate purposes. To avoid implications concerning the behavior of the various curves where m is very small, the slopes of the apparent tangents of the curves as $m \rightarrow 0$ might appropriately be termed "pseudo-limiting slopes." The only other strong acids for which data are included in the *International Critical Tables* are HBr and H_2SO_4 . The curve for HBr is similar to those for HCl and HNO_3 . The behavior of H_2SO_4 is clearly exceptional.

There is an interesting relation between the various curves of FIGURE 3, namely, the difference between the slopes of the NaCl and NaNO_3 curves is nearly the same as the difference between the slopes of the KCl and KNO_3 curves. The relationship between the values of $\Delta\gamma$ characteristic of these four binary solutions of molality, m' is

$$\Delta\gamma(\text{NaCl}, m') - \Delta\gamma(\text{NaNO}_3, m') = \Delta\gamma(\text{KCl}, m') - \Delta\gamma(\text{KNO}_3, m'). \quad (3)$$

Although the $\Delta\gamma$ - m relationship for the acids is not linear, an equation similar to EQUATION 3 applies. At any molality m'

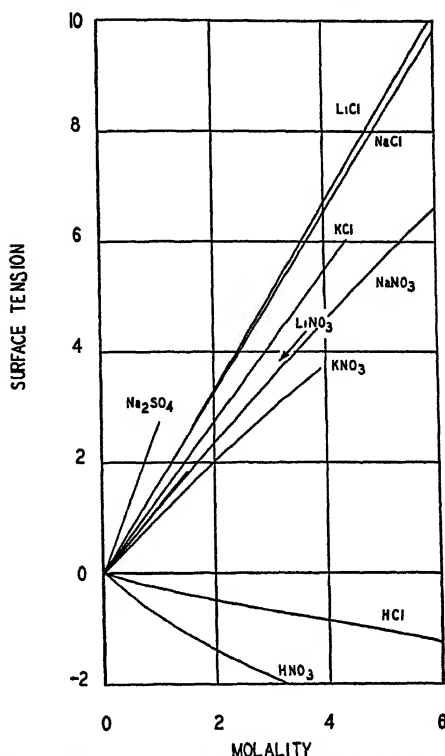


FIGURE 3 Surface tension increments, $\Delta\gamma$, (dyne/cm) at 20°C of aqueous solutions of typical electrolytes.

$$\Delta\gamma(\text{NaCl}, m') - \Delta\gamma(\text{NaNO}_3, m') = \Delta\gamma(\text{HCl}, m') - \Delta\gamma(\text{HNO}_3, m'). \quad (4)$$

These relationships may be summarized in the statement that the contributions of the various ions to γ , or to $\Delta\gamma$, are additive when there are but two species of ions in solution. For an electrolyte having the formula $M_{r+}X_{r-}$ in a solution of molality, m'

$$\Delta\gamma(M_{r+}X_{r-}, m') = \sum_i \Delta\gamma(i, v, m'). \quad (5)$$

Each $\Delta\gamma_i$ in the sum must have the value characteristic of the molality, v, m' of the i^{th} ion in the m' molal solution of $M_{r+}X_{r-}$. Whenever EQUATION 2 is valid, it may be combined with EQUATION 5. The surface tension of a mixture of electrolytes may then be computed from the $\Delta\gamma$ values characteristic of the constituent ions.

Methods for the Calculation of the Surface Tensions of the Hypothetical Solutes $\text{H}\cdot\text{H}\cdot\text{SO}_4$ and $\text{H}\cdot\text{HSO}_4$. With the aid of EQUATION 5, the surface tensions of solutions containing only a single hypothetical solute may be computed. That containing the solute $\text{H}\cdot\text{H}\cdot\text{SO}_4$ at the molality m' is given by

$$\Delta\gamma(\text{H}\cdot\text{H}\cdot\text{SO}_4, m') = \Delta\gamma(\text{HCl}, 2m') + \Delta\gamma(\text{Na}_2\text{SO}_4, m') - \Delta\gamma(\text{NaCl}, 2m'). \quad (6)$$

Since each mole of $\text{H}\cdot\text{H}\cdot\text{SO}_4$ consists of one mole of SO_4^{--} and two moles of H^+ , the molality of HCl and NaCl from which the respective values $\Delta\gamma$ are to be determined must have the value $2m'$. This fact is indicated in the equations by insertion of $2m'$ to indicate m of the respective electrolyte. The surface tensions of Na_2SO_4 solutions may be represented sufficiently well over the range of concentration and temperature investigated by the simple equation

$$\Delta\gamma(\text{Na}_2\text{SO}_4, m) = 2.73 m = 2.73 m'. \quad (7)$$

The surface tension of NaCl solutions may likewise be represented by

$$\Delta\gamma(\text{NaCl}, m = 2m') = 1.64 m = 3.28 m'. \quad (8)$$

EQUATIONS 7 and 8 may be substituted in EQUATION 6 to yield a simpler equation for $\text{H}\cdot\text{H}\cdot\text{SO}_4$

$$\Delta\gamma(\text{H}\cdot\text{H}\cdot\text{SO}_4, m') = \Delta\gamma(\text{HCl}, 2m') - 0.55 m'. \quad (9)$$

In a similar manner, an equation may be written for the surface tension increment produced by the hypothetical solute $\text{H}\cdot\text{HSO}_4$

$$\Delta\gamma(\text{H}\cdot\text{HSO}_4, m') = \Delta\gamma(\text{HCl}, m') + \Delta\gamma(\text{Na}\cdot\text{HSO}_4, m') - \Delta\gamma(\text{NaCl}, m'). \quad (10)$$

The use of EQUATION 10 depends upon information concerning $\Delta\gamma$ of the hypothetical solute $\text{Na}\cdot\text{HSO}_4$ whose solutions contain only the two ions Na^+ and HSO_4^- . Since the HSO_4^- dissociates in all real solutions of NaHSO_4 , values of $\Delta\gamma$ cannot be determined by direct experiment. However, it is reasonable to assume that $\Delta\gamma$ of $\text{Na}\cdot\text{HSO}_4$, like that of other salts, is proportional to m and nearly independent of temperature. With the aid of that assumption, an equation similar to EQUATION 9 may be written for $\text{H}\cdot\text{HSO}_4$.

$$\Delta\gamma(\text{H}\cdot\text{HSO}_4, m') = \Delta\gamma(\text{HCl}, m') + m'\Gamma. \quad (11)$$

The quantity Γ , which is defined by

$$m'\Gamma = \Delta\gamma(\text{Na}\cdot\text{HSO}_4, m') - \Delta\gamma(\text{NaCl}, m'), \quad (12)$$

was determined empirically from the variations with composition of $\Delta\gamma$ of H_2SO_4 in solutions in which the concentrations of $\text{H}\cdot\text{H}\cdot\text{SO}_4$ and $\cdot\text{HHSO}_4$ are relatively small. The assumptions just mentioned lead to the conclusion that Γ is independent of temperature and of molality.

The contribution of the undissociated $\cdot\text{HHSO}_4$ to the surface tensions of H_2SO_4 solutions can not be calculated at the present time. Fortunately, its concentration is small in the solutions of greatest interest. The effect of $\cdot\text{HHSO}_4$ in concentrated solutions will be discussed qualitatively, below.

A preliminary attempt to explain the anomalies in the curves of FIGURE 1

was unsuccessful. In the absence of adequate data for the surface tension of hydrochloric acid solutions at temperatures above and below 20°, the assumption was made that $\Delta\gamma$ could be treated as independent of temperature. The resultant estimate of the depth of the minimum in the surface tension curve for 0° was much too small; the calculated $\Delta\gamma$ was less than 10 per cent of the observed value. An estimate of the molality at which the minimum appeared was also seriously in error.

Although several assumptions and approximate empirical relationships were involved in that calculation, the only one which seemed likely to be seriously at fault was the assumption of the temperature independence of $\Delta\gamma$ of HCl solutions. Nearly all of the measurements upon which the tabulations in the *International Critical Tables* were based had been made at 18° to 20°C. A series of values had been reported¹⁰ for 16°, but these were not internally consistent and were useless in an investigation of the effect of temperature on surface tension. There were also two isolated measurements⁴ of one molal solution at 0° and 30°C., but they did not appear to be reliable and it has since been shown that those two values of $\Delta\gamma$ are both low. The difference between them, however, strengthened the suspicion that $\Delta\gamma$ of HCl solutions becomes more negative as temperature is lowered. In addition, recent measurements made at 25° by Belton⁸ led to $\Delta\gamma$ values about half as large as those tabulated for 20°C. Since the best 18° to 20° work had been done well over fifty years ago, it was necessary that measurements be made at several temperatures, by one method, on the same set of solutions.

Surface Tensions of Aqueous Hydrochloric Acid Solutions. Three new series of measurements were made in this laboratory by Young, Runge, and Grinstead.¹¹ The method employed was the drop weight procedure developed by Harkins and Brown.³ It was shown that the weights of drops falling from sharp, carefully ground tips are accurately reproducible when each drop is carefully enlarged, just prior to its fall, under conditions closely approaching a static state. From the weights, precise surface tension values may be computed for liquids varying widely in viscosity, density, surface tension, etc.

The new data are shown in FIGURE 4. The radii of the circles correspond to a precision of 0.1 per cent in γ . The agreement (to about 0.05 per cent) of a single measurement with the 25° bubble pressure data of Belton⁸ is excellent. Moreover, the old *International Critical Table* curve for 20° is less than 0.1 per cent lower at 5 molal and only about 0.25 per cent lower at 8.4 molal than the new curve. This agreement is in part a fortuitous result of the curve which Young and Harkins⁴ drew through the points representing the older work. The uncertainties in temperatures alone in the very old work could account for the small discrepancies observed. The general agreement of the three methods—capillary height, bubble pressure,

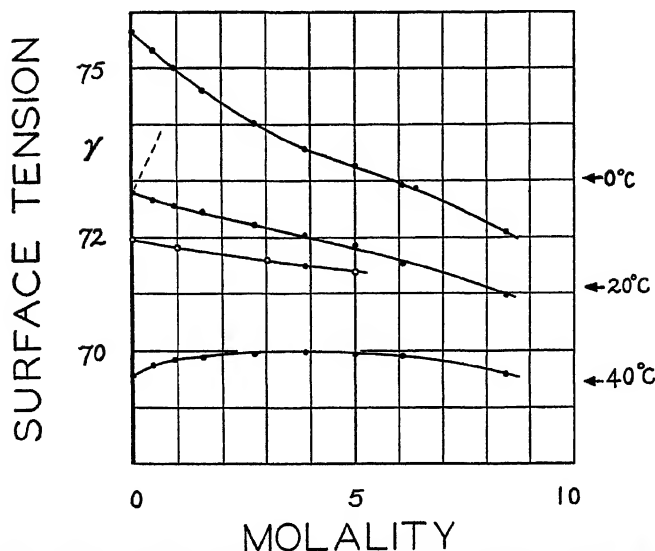


FIGURE 4. Surface tensions (dyne/cm.) of aqueous HCl solutions. Dashed line shows surface tensions of NaCl solutions, for comparison.

and drop weight—and the reproducibility of the solutions prepared throughout a period of seventy years is entirely satisfactory.

The rapid variation of $\Delta\gamma$ with temperature is remarkable in view of the fact that the variations in the increments due to other electrolytes are smaller than experimental error. At 0°C., $\Delta\gamma$ of dilute HCl solutions is about five times the corresponding increments at 25°C. At 40°C., $\Delta\gamma$ is actually positive for all values of m less than 8.5 molal. At 0°C., where the effects are largest, the curve is nearly linear.

Measurements were made only at 0°, 20°, and 40°. Values for 18° are needed for the proposed calculations. These were interpolated by means of quadratic equations and are shown in TABLE 1. A single value due to

TABLE 1
 $\Delta\gamma$ OF HCl SOLUTIONS AT 18°C.

m	0.45	0.92	1.56	2.73	3.87	5.00	6.08	8.44	17.7
$\Delta\gamma$	-.16 _s	-.27 _s	-.42	-.67	-.89	-1.08	-1.41	-2.00	-6.8*

* From Quincke.

to Quincke¹² for a 17.7 molal solution was taken from the 20° table of the *International Critical Tables*.

Empirical Equations for the Calculation of $\Delta\gamma$ of $\text{H}\cdot\text{H}\cdot\text{SO}_4$ and $\text{H}\cdot\text{HSO}_4$. Sufficient data were now available for the completion of the calculations

of $\Delta\gamma$ of $\text{H}\cdot\text{H}\cdot\text{SO}_4$. To obtain $\Delta\gamma$ of an $\text{H}\cdot\text{H}\cdot\text{SO}_4$ solution of molality m' , values of $\Delta\gamma$ for HCl of molality $2m'$ were substituted in EQUATION 9. Since Nw , the weight normality,¹¹ is the same for these two solutions, it is convenient to compare them at selected values of Nw . Column 1, in TABLE 2,

TABLE 2
 $\Delta\gamma$ OF AQUEOUS SOLUTIONS AT 0°C.

(1)	(2)	(3)	(4)	(5)	(6)
Nw^*	HCl	$\text{H}\cdot\text{H}\cdot\text{SO}_4$	$\text{H}\cdot\text{HSO}_4$	$\text{Na}\cdot\text{HSO}_4$	m of $\text{H}\cdot\text{H}\cdot\text{SO}_4$
.45	-0.32	-0.44	+0.085	+1.14	0.225
.92	-.63	-.88	.20	2.34	.46
1.56	-1.04	-1.47	.36	3.96	.78
2.73	-1.62	-2.37	.84	6.93	1.36 _s

* Nw is the molality for all solutes except $\text{H}\cdot\text{H}\cdot\text{SO}_4$.

contains values of Nw , column 2 contains $\Delta\gamma$'s of HCl and column 3 shows the newly calculated $\Delta\gamma$'s for $\text{H}\cdot\text{H}\cdot\text{SO}_4$. The corresponding molalities of $\text{H}\cdot\text{H}\cdot\text{SO}_4$ are in column 6. Similar calculations were made for 18°, 20°, 30°, and 40°. Those for 18° will be needed in our subsequent calculations and are shown in columns 1, 2, 3, and 6 of TABLE 3.

TABLE 3
 $\Delta\gamma$ OF AQUEOUS SOLUTIONS AT 18°C.

(1)	(2)	(3)	(4)	(5)	(6)
Nw	HCl	$\text{H}\cdot\text{H}\cdot\text{SO}_4$	$\text{H}\cdot\text{HSO}_4$	$\text{Na}\cdot\text{HSO}_4$	m of $\text{H}\cdot\text{H}\cdot\text{SO}_4$
.45	-0.16 _s	-0.29	+0.15	+1.14	0.225
.92	-.27 _s	-.53	.37	2.34	.46
1.56	-.42	-.85	.67	3.96	.78
2.73	-.67	-1.42	1.24	6.93	1.36

$\Delta\gamma$ of the hypothetical solute $\text{H}\cdot\text{HSO}_4$ was then computed in a similar manner by means of EQUATION 11. The results are in column 4. For comparison, $\Delta\gamma$ calculated for hypothetical $\text{Na}\cdot\text{HSO}_4$ is given in column 5. Both of these calculations were based upon a value for Γ (cf. EQUATIONS 11 and 12) of 0.9. This value was determined from the slopes of the experimental $\Delta\gamma$ - m curves in concentration regions in which $\text{H}\cdot\text{HSO}_4$ seems to be the principal solute.

The First Test of the Theory of the Anomalous Surface Tensions of H_2SO_4 Solutions. The most significant test to be applied to the proposed explanation of the anomalies in the curves of FIGURE 1 is a comparison of the curves calculated for $\text{H}\cdot\text{H}\cdot\text{SO}_4$ with the experimental curves for H_2SO_4 . The

* Equivalents of solute per kilogram of solvent.

comparison is made in FIGURE 5, where the heavy lines represent experimental values of γ of H_2SO_4 solutions and the short lighter lines the curves

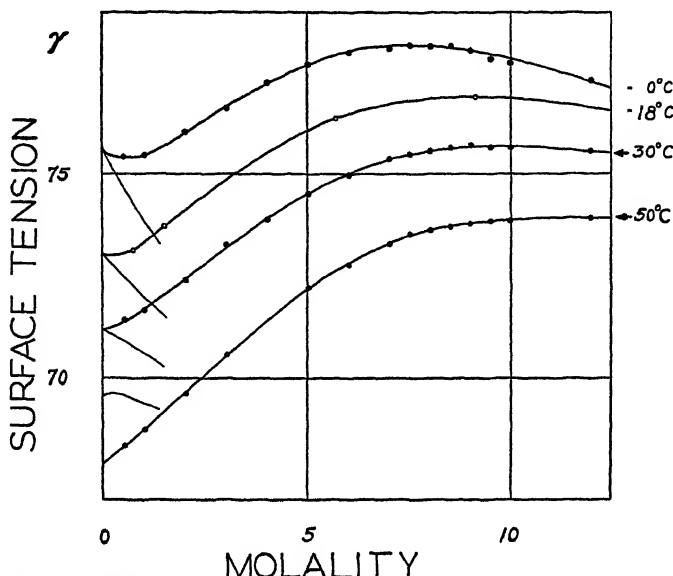


FIGURE 5. Surface tension (dyne/cm.) versus m (mole per kilogram of solvent.) The four heavy lines are experimental curves for H_2SO_4 at 0°, 18°, 30°, and 50°C. The symbol \circ denotes data of Rontgen and Schneider; \bullet denotes surface tensions calculated from measurements of Morgan and Davis.

The short lines are calculated isotherms for the hypothetical solute $\text{H}\cdot\text{H}\cdot\text{SO}_4$ at 0°, 18°, 30°, and 40°C.

calculated for $\text{H}\cdot\text{H}\cdot\text{SO}_4$. At each temperature, α_2 , the degree of dissociation of the HSO_4^- increases toward unity as molality is reduced. At sufficiently low concentrations, therefore, the slope of the experimental curve should approach that of the calculated curve. The approach should be more rapid at low temperatures, because the second dissociation constant, K_2 , of sulfuric acid increases as temperature is lowered.^{7, 13} At 0°C., where K_2 is relatively large, and where $\Delta\gamma$ of $\text{H}\cdot\text{H}\cdot\text{SO}_4$ is large and negative, the experimental curve exhibits a prominent minimum, as it must if it is to become tangent to the theoretical curve. At 50°C., however, the calculated values of $\Delta\gamma$ are all positive. There is, therefore, no reason to expect that the experimental curve should possess a minimum, and there is no indication in the data that it does. A point of inflection, however, is to be expected and it is apparent in the experimental curve. At 18°, the comparison of calculated curves with experiment is more difficult, because 18°C. is close to, though below, the temperature at which the "pseudo-limiting slope" of the $\text{H}\cdot\text{H}\cdot\text{SO}_4$ curve changes sign. A minimum is to be expected, therefore, but it should be almost undiscernible and close to the molality axis. At slightly higher temperatures, as $\Delta\gamma$ of $\text{H}\cdot\text{H}\cdot\text{SO}_4$ changes sign, the minimum should disappear entirely. The data for 18°C. and 30°C. are in accord with these

expectations. They are not adequate, however, to determine the existence or absence of a shallow minimum at 18°C.

The theory can be tested further by a calculation of the position and depth of the minima at 0° and at 18°. For that purpose, it is necessary to estimate concentrations of SO_4^{--} and HSO_4^- . The second dissociation constant of H_2SO_4 has been measured between 5° and 55°C. in a recent investigation of Young, Klotz, and Singleterry⁷ (*cf.* Harned and Owen¹³, pp. 429 and 580.) From their equations, values of K_2 at 0° and at 18° were computed. They are .0201 and .0124, respectively. For the computation of ion concentrations, it is also necessary to estimate the quotient γ_R in the equilibrium equation

$$\frac{(1 + \alpha_2)(\alpha_2)}{(1 - \alpha_2)} M\gamma_R = K_2. \quad (13)$$

Here, M represents the stoichiometric molality of H_2SO_4 minus the molality of the undissociated acid, $\cdot\text{HHSO}_4$. In all except very concentrated solutions, M and m are nearly equal and may often be used interchangeably. γ_R represents the product of the activity coefficients of H^+ and SO_4^{--} divided by the activity coefficient of HSO_4^- . The symbol γ will be used in association with activity coefficients only when it carries the subscript R . The subscript will avoid confusion between the two roles customarily assigned to this single symbol.

The Activity Coefficient Ratio:

THEORETICAL CALCULATIONS. For very dilute solutions, the activity coefficients of each ion may be computed approximately as Kielland¹⁴ has done from the Debye-Hückel¹⁵ theory and reasonable values of the parameter, a .

CONDUCTIVITY DATA. For somewhat more concentrated solutions, a small table of values of γ_R has been calculated by Bray and Liebhafsky.¹⁶ They based their calculations on the theoretical analysis made by Sherrill and Noyes⁵ of the conductivity data of Noyes and Stewart.¹⁷

INTENSITIES OF RAMAN LINES. For the most concentrated solutions, γ_R may be computed from the intensities of three Raman lines studied by Rao.¹⁸ Rao ascribed the lines of 980, 1043, and 910 wave numbers to SO_4^{--} , HSO_4^- , and undissociated sulfuric acid, respectively. He measured the relative intensities of each line produced in sulfuric acid solutions of various concentrations. He also measured intensities of the line of 980 cm^{-1} arising from 3 molar $(\text{NH}_4)_2\text{SO}_4$ and 3 molar KHSO_4 and of the 1043 cm^{-1} line from 3 molar KHSO_4 . Rao's calculations were incorrectly made (*cf.* Redlich¹⁹), but from the tabulated intensity data, concentrations of undissociated sulfuric acid and the concentrations of each of the ions were computed. The concentrations and intensities are shown in TABLE 4 and in FIGURE 6. Rao did not report the temperature of his solutions, but stated

TABLE 4
RELATIVE INTENSITIES OF RAMAN LINES AND CONCENTRATIONS CALCULATED FROM THEM

(1)	(2)	(3)	(4)	(5)	(6)	(7)
C H_2SO_4	I 980 cm^{-1}	C SO_4^{2-}	I 1043 cm^{-1}	C HSO_4^-	I 910 cm^{-1}	C $\cdot HHSO_4$
16.9			860	1.28	2330	15.6
14.9			1800	2.68	1700	12.2
12.8			3120	4.64	1150	8.2
10.8	11.8	.25	4350	6.48	765	4.07
8.7	15.2	.32 ₄	4412	6.57		1.81
6.8	13.0	.27 ₇	3678	5.49		1.03
3.0	6.5	.13 ₈	1800	2.68		.18
$(NH_4)_2 SO_4$ 3.0	140.5	3.0				
$KHSO_4$ 3.0	8.7	.18 ₅	1890	2.81 ₅	8.7	

(1) Stoichiometric Concentrations (mole/liter of solution) of electrolytes; (2), (4), and (6) relative intensities I of Raman lines; (3), (5) concentrations, c , of ions; (7) concentration of undissociated sulfuric acid ($\cdot HHSO_4$) computed by difference.

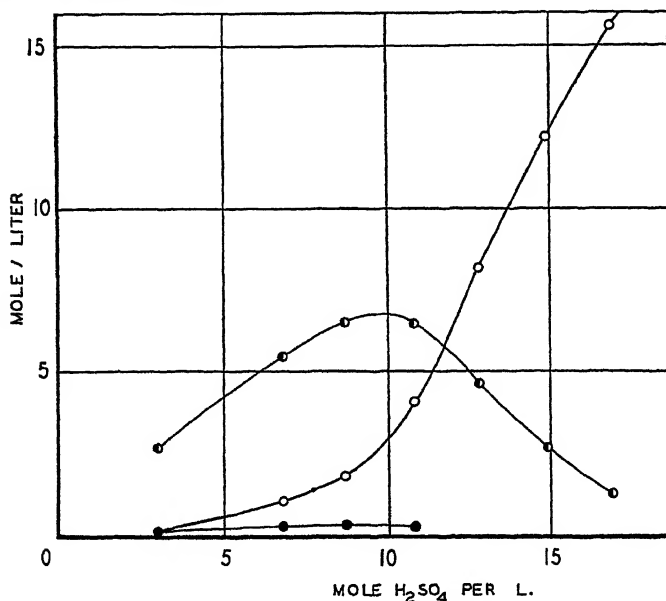


FIGURE 6. Concentrations, calculated from Rao's Raman intensity data of constituents of sulfuric acid solutions: ●, concentration of $H\cdot H\cdot SO_4$; ●, concentration of $H\cdot HSO_4$; ○, concentration of $\cdot HHSO_4$.

that they were cooled by water during exposure. In the absence of information concerning the temperature we have taken it to be 25° , since calculations were somewhat simplified by that assumption.

Each of the concentrations (in moles per liter of solution) was converted

to a molality (moles per kilogram of solvent), by means of the appropriate density data. The molalities so computed, and the dissociation constant, K_2 , determined by Young, Klotz, and Singleterry⁷ were substituted in EQUATION 13 and γ_R calculated. γ_R from the various sources was then plotted against the square root of the ionic strength μ . The values tabulated for more dilute solutions by Bray and Liebhafsky were plotted on the same graph. One line was then drawn to represent γ_R as a function of $\sqrt{\mu}$ up to $\sqrt{\mu} = 3.7$.

Calculation of Degree of Dissociation of HSO_4^- at 0°C. and 18°C. The γ_R isotherm is presumably valid for a temperature in the neighborhood of 25°C. For the calculations to be made, the activity coefficient ratio is required for 0°C. and also for 18°C. Tables of activity coefficients show that, as temperature is lowered, the activity coefficients of HCl (in the concentration ranges of greatest interest here) increase, and that the activity coefficients of alkali metal sulfates decrease (*cf.* Harned and Owen,¹⁸ pp. 415 and 547). Unless there is a very rapid change due to the bisulfate ion, γ_R probably does not change rapidly with temperature. Some evidence that the rate of change of γ_R with temperature is small lies in the agreement between the two segments of the γ_R curve. One portion is for 25° and the other (from Raman data) is probably for a somewhat different temperature. The single γ_R curve was used, in the absence of adequate data, as a standard curve for all temperatures between 0 and 25°C.

For use in the calculations described below, curves were constructed showing α_2 at 0°C. as functions of the molality, m . The points were calculated from the standard γ_R curve and the appropriate values of K_2 by a method of successive approximations. (For a selected value of α ; m , μ , and γ_R were obtained, in that order. Substitution of γ_R in EQUATION 13 yielded the next approximation of m .)

Additional Tests of the Dissociation Theory of the Surface Tensions of H_2SO_4 Solutions. Methods have now been made available for a calculation of theoretical points on the surface tension isotherms. Because of inadequacies in the activity coefficient data and in the surface tension data themselves, great precision can not be expected. It is interesting, nevertheless, to ascertain how well the calculated and experimental curves will agree. Theoretical calculations of $\Delta\gamma$ were completed as in the following example: For a 0.7 molal solution, $\alpha_2 = 0.18$ at 0°, and 0.12 at 18°C. The molalities of SO_4^{--} are therefore 0.12 and 0.08, respectively. The contributions of $\text{H}\cdot\text{H}\cdot\text{SO}_4$ to $\Delta\gamma$ were read, for these molalities, from theoretical curves such as those in FIGURE 5. They are -0.25 and -0.12 . From similar plots of data in TABLES 2 and 3, the contributions of the $\text{H}\cdot\text{HSO}_4$ were determined to be $+0.11$ for 0° and $+0.38$ dyne/cm. at 18°C. The results of a series of such calculations are summarized in TABLE 5, where they are compared with available experimental data.

The discrepancies in TABLE 5 are within the limits of precision of the experimental data (surface tensions, activity coefficients, etc.) The results

TABLE 5
COMPARISON OF CALCULATED AND EXPERIMENTAL ISOTHERMS

	0°C		18°C	
	Calculated	Observed	Calculated	Observed
Molality at which $\Delta\gamma = 0$	1.5	1.3	.04	small*
Molality at which $d\gamma/dm = 0$	0.65	0.5 to 0.7	.02	small*
Minimum $\Delta\gamma^\circ$	-0.15	-0.21	-.01	small*
$\Delta\gamma$ at $m = 0.7$	-0.15	-0.21	+.25	—

* Too small to estimate.

argue, therefore, for the validity of the analysis of the problem presented here. If future experimental work should show that the value of $\Delta\gamma$ at 0° is as large as -0.21, the explanation probably is that the calculated value of α_2 is too small, because the estimated value of γ_R is too large for 0°C. More extensive and more precise Raman density data are very desirable for the further elucidation of such problems as these.

There is one other test which can be applied to the theory: position and height of the maximum surface tension at 18°C. may be computed.* Because some of the assumptions and approximations used in the previous calculations are not likely to be valid at such high concentrations, little more than the orders of magnitude can be estimated. Probably the most serious source of error is the lowering in surface tension due to the presence of undissociated sulfuric acid, $\cdot\text{HHSO}_4$. The solvent in which the ions are dissolved may be regarded as a mixed solvent, consisting of H_2O and $\cdot\text{HHSO}_4$. The surface tension curves for mixtures of water and undissociated liquids such as alcohols, esters, organic acids, *etc.*, usually lie well below straight lines drawn on mole fraction plots to connect the surface tensions of the pairs of undissociated liquids. In a 65% solution of H_2SO_4 , the molality of $\cdot\text{HHSO}_4$ is about 6.5, and is increasing rapidly as concentration increases. The maximum surface tension should, therefore, be reached in a solution considerably more dilute than the one in which the molality of HSO_4^- reaches a maximum (65%). Actually, the maximum surface tension is reached in a 47 weight per cent solution. The comparison is satisfactory.

When the molality of HSO_4^- is 7.3, the contribution of $\text{H}\cdot\text{HSO}_4$ to $\Delta\gamma$ taken from an extension of TABLE 3 is 5.0 dyne/cm. Since the mole fraction of $\cdot\text{HHSO}_4$ in the mixed solvent (H_2O and $\cdot\text{HHSO}_4$) is .024 and since the

* A different explanation of the maximum has been proposed by SABININA AND TERPUGOW.³⁰ They ascribed it to the formation of a hydrate. The behavior of very concentrated solutions, especially those more concentrated than 80 weight per cent, is undoubtedly complicated by the formation of solvates. BIRON³¹ has presented strong evidence for the existence of the monohydrate of H_2SO_4 (*cf.* Lewis and Randall³²). In view of all the evidence presented here, that large concentrations of ions exist, the explanation offered by Sabinina and Terpugow seems to be untenable.

difference between the surface tension of water and pure H_2SO_4 is about 21 dyne/cm., the minimum correction to be applied to the calculated surface tension is 0.5 dyne/cm.; it is probably more nearly twice 0.5 (compare surface tension data for many binary aqueous solutions in the *International Critical Tables*¹). The difference between the experimental 3.9 and the uncorrected 5.0 dynes per cm. at 47 per cent is, therefore, about what is to be expected.

The satisfactory correlation of surface tensions and Raman intensities becomes even more interesting when those data are compared with the conductivities of sulfuric acid solutions. In Kohlrausch's²⁸ curve for 18°C., the specific conductance reaches a maximum in a 30 weight per cent solution. Because of the likelihood of rapidly changing ionic mobilities, the difference between the concentrations of the solutions in which the specific conductance is a maximum and in which ion concentrations reach a maximum is not too large. Indeed, the product of the specific conductance and the viscosity attains a maximum in a solution having much more nearly the concentration in which the H^+ and HSO_4^- concentrations are largest (FIGURE 7). It is

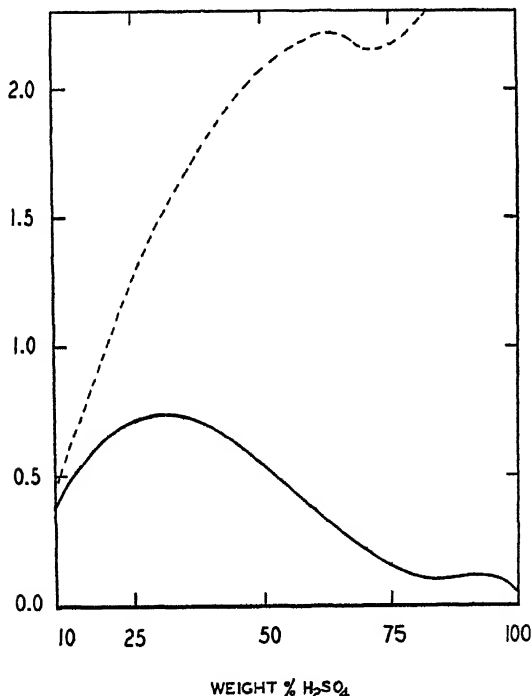


FIGURE 7 Specific conductance ($\text{ohm}^{-1} \text{cm.}^{-1}$) of aqueous sulfuric acid solutions. Light dashed curve represents product of specific conductance and specific viscosity.

interesting to note also that the concentration in which the maximum specific conductance is observed has been shown by Vinal and Craig²⁴ to move to

higher concentrations as temperature rises. So, too, does the maximum in the surface tension curves. All of these properties are satisfactorily correlated by the simple dissociation theory even in solutions so concentrated as 50 and 60 weight per cent acid.

Summary

The anomalies in the surface tension composition isotherms of sulfuric acid have been explained in terms of electrolytic dissociation. The depth and position of the minimum in the zero degree isotherm were computed from considerations of the opposing effects of sulfate and hydrogen ions on the one hand and hydrosulfate and hydrogen ions on the other. Calculated changes in the quantitative effects of the respective ions account for the shifting (toward $m = 0$) and gradual disappearance of the minimum (at about 35°), as the temperature rises. The maximum $\Delta\gamma$ which is prominent at all temperatures seems to be due principally to the maximum concentration attained by the bisulfate ion (and hydrogen ion) as concentration increases. The magnitude and position of the surface tension maximum can be computed only roughly because of the complicating contribution of undissociated sulfuric acid, but the agreement is satisfactory.

The explanations of both the minimum and the maximum are in accord with other data including specific conductance and the intensities of Raman lines.

Bibliography

1. RÖNTGEN, W. C., & J. SCHNEIDER. 1886. *Ann. Phys.* (3) **28**: 165.
2. MORGAN, J. L. R., & C. E. DAVIS. 1916. *J. Am. Chem. Soc.* **38**: 555.
3. HARKINS, W. D., & F. E. BROWN. 1919. *J. Am. Chem. Soc.* **41**: 499.
4. YOUNG, T. F., & W. D. HARKINS. 1928. *International Critical Tables of Numerical Data, Physics, Chemistry and Technology* IV: 463. McGraw-Hill. New York.
5. SHERRILL, M. S., & A. A. NOYES. 1926. *J. Am. Chem. Soc.* **48**: 1861.
6. KLOTZ, I. M., & C. F. ECKERT. 1942. *J. Am. Chem. Soc.* **64**: 1878.
7. YOUNG, T. F., I. M. KLOTZ, & C. R. SINGLETERRY. Unpublished.
8. BELTON, J. W. 1935. *Trans. Far. Soc.* **31**: 1413.
9. ONSAGER, L., & N. N. T. SAMARAS. 1934. *J. Chem. Phys.* **2**: 528.
10. WHATMOUGH, W. H. 1901. *Z. phys. Chem.* **39**: 129.
11. YOUNG, T. F., P. W. RUNGE & S. R. GRINSTEAD. 1948. Unpublished.
12. QUINCKE, G. 1877. *Ann. Phys.* (2) **160**: 337.
13. HARNED, H. S., & B. B. OWEN. 1943. *The Physical Chemistry of Electrolytic Solutions*. Reinhold. New York.
14. KIELLAND, J. 1937. *J. Am. Chem. Soc.* **59**: 1675.
15. DEBYE, P., & E. HÜCKEL. 1923. *Phys. Z.* **24**: 185.
16. BRAY, W. C. & H. A. LIEBHAFSKY. 1935. *J. Am. Chem. Soc.* **57**: 51.
17. NOYES, A. A. & M. A. STEWART. 1910. *J. Am. Chem. Soc.* **32**: 1133.
18. RAO, N. R. 1940. *Ind. J. Phys.* **14**: 143.
19. REDLICH, O. 1946. *Chem. Rev.* **39**: 333.
20. SABININA, L., & L. TERPUGOW. 1935. *Z. phys. Chem.* **173**: (A): 237.
21. BIRON, E. von. 1899. *J. Russ. Phys. Chem. Soc.* **31**: 190.
22. LEWIS, G. N. & M. RANDALL. 1923. *Thermodynamics and the Free Energy of Chemical Substances*: 87. McGraw-Hill. New York.
23. KOHLRAUSCH, F. 1856. *Ann. Phys.* (2) **159**: 233.
24. VINAL, G. W., & D. N. CRAIG. 1934. *J. Res. N. B. S.* **13**: 689.

THE DIFFUSION COEFFICIENT OF POTASSIUM CHLORIDE IN AQUEOUS SOLUTION AT 25°C.

By HERBERT S. HARNED AND RALPH L. NUTTALL

Department of Chemistry, Yale University, New Haven, Connecticut

In November, 1944, at a conference on "The Diffusion of Electrolytes and Macromolecules in Solution"¹ under the auspices of the New York Academy of Sciences, Harned and French² described a conductance method for the determination of the diffusion coefficients of electrolytes in dilute solutions. Four determinations of the diffusion coefficient of potassium chloride in water at 25°C. in the concentration range, 0.0025 to 0.005 molal, were made. The error of these determinations was estimated to be $\pm 0.9\%$, but Harned and French concluded by stating that "our experience indicates that considerable improvement in accuracy may be effected."

By improving the technique for the elimination of turbulent flow and temperature gradients, we have succeeded in obtaining an accuracy of the order of ± 0.05 to $\pm 0.2\%$. The recent apparatus and technique have been described in detail.³ Our results for potassium chloride in the concentration range, 0.001 to 0.01 molal, confirm the limiting value for the diffusion coefficient as derived from the Nernst equation and are in excellent agreement with the theory of Onsager and Fuoss.⁴

Determinations of the diffusion coefficient of potassium chloride in water at 25°, extended to include the concentration range from 0.001 to 0.526 molal, will be presented here. These results are of sufficient accuracy to test the theory for this typical 1-1 electrolyte.

The Theory of Electrolytic Diffusion⁴

For an electrolyte diffusing in the x direction, Fick's first law may be expressed by

$$J = n\mathbf{v} = -\mathcal{D} \frac{\partial n}{\partial x} = -\mathfrak{M} \frac{\partial \mu}{\partial x} \quad (1)$$

where J is the flow, n the number of molecules of diffusing component per cc., \mathbf{v} its velocity, and \mathcal{D} its diffusion coefficient. The form of Fick's law given by the last term on the right of this equation, which recognizes that "the force" causing diffusion is the gradient of the chemical potential of the diffusing substance, was first suggested by Hartley⁵ and is a necessary basis for all subsequent theoretical computations. It is clear that

$$\mathcal{D} = \left(\frac{\partial \mu}{\partial n} \right)_{P,T} \mathfrak{M}. \quad (2)$$

As defined above, the flow is the number of molecules per cc. times the velocity. To convert to molar units, we let

$$n = N\bar{n}; \quad \mu N = \bar{\mu} \quad (3)$$

where N is Avogadro's number and \bar{n} is in mols per cc. The flow in mols per cc. per second is given by

$$\bar{n}v = -\bar{\mathcal{D}}\bar{n} \frac{\partial \bar{\mu}}{\partial x} = -\left(\frac{\bar{\mathcal{D}}\bar{n}}{N^2}\right) \frac{\partial \bar{\mu}}{\partial x} \quad (4)$$

and, if c is the concentration in mols per liter, then

$$\mathcal{D} = \frac{\partial \bar{\mu}}{\partial \bar{n}} \bar{\mathcal{D}}\bar{n} = 1000 \bar{\mathcal{D}}\bar{n} \frac{\partial \bar{\mu}}{\partial c}. \quad (5)$$

Upon introducing the equation which defines the activity coefficient of an electrolyte on the mols per liter scale,

$$\mu = \mu^0 + \nu RT \ln \gamma_{\pm} c_{\pm} \quad (6)$$

we readily obtain the equation

$$\mathcal{D} = \nu 1000 RT \frac{\bar{\mathcal{D}}\bar{n}}{c} \left(1 + c \frac{\partial \ln \gamma_{\pm}}{\partial c}\right). \quad (7)$$

At the limit, when c equals zero, the term in parenthesis becomes unity, $\frac{\bar{\mathcal{D}}\bar{n}}{c}$ is a constant containing the limiting ionic conductances and the equation reduces to the limiting formula of Nernst. The major contribution to the variation of the diffusion coefficient with the concentration is accounted for by the thermodynamic term in the parenthesis. However, the theory of the effect of the interactions between the ions and water molecules indicates that $\bar{\mathcal{D}}\bar{n}/c$ is a function of the concentration. This theory is described in detail in other places^{4, 6, 7} and only a brief outline and a statement of the final result is necessary for the present purpose.

The theory is simpler than that of conductance because of the fact that electrical neutrality must be maintained and, consequently, both ions diffuse with the same velocity. This eliminates the "time of relaxation effect" whose terms contain the difference in the two velocities. There remains the effect of electrophoresis, first introduced into the theory of electrolytes by Debye and Hückel,⁸ and applied in an extended form to the theory of electrolytic diffusion by Onsager and Fuoss.⁴

The term electrophoresis has been used to denote the effect of the relative motion of the ions in respect to the solvent molecules. According to the theory of Debye, each ion is surrounded with an atmosphere of ions of opposite charge equal to that of the ion. If, as in the case of conductance, the ion is subjected to an electrical field, it will migrate, but at the same time its atmosphere and the solvent associated with it will move in the opposite direction. This counterwise motion has the effect of retarding the velocity of the ion.

In diffusion, where both ions migrate with the same velocity, electrophoresis occurs because, as the ions move in one direction, they replace solvent molecules which move in the opposite direction. The electrophoretic effect was computed by Onsager and Fuoss by evaluating the net force on an element of volume at a distance r from an ion, and substituting this force in Stokes' law for the motion of a sphere in a viscous fluid. This required the Boltzmann equation for the number of i -ions in the presence of a j -ion and Debye and Hückel's value for the potential of the j -ion and its atmosphere.

The final result for the diffusion coefficient of a 1-1 type electrolyte is given by

$$\mathcal{D} = 16.629 \times 10^{10} T \frac{\overline{\mathcal{D}}}{c} \left(1 + c \frac{\partial \ln y_{\pm}}{\partial c} \right) \quad (8)^*$$

where

$$\begin{aligned} \left(\frac{\overline{\mathcal{D}}}{c} \right) 10^{20} = & 1.0748 \left(\frac{\lambda_1^0 \lambda_2^0}{\Lambda^0} \right) - \frac{22.148}{\eta_0 (DT)^{\frac{1}{2}}} \left(\frac{\lambda_1^0 - \lambda_2^0}{\Lambda^0} \right)^2 \frac{\sqrt{c}}{1 + A' \sqrt{c}} \\ & + \frac{9.304 \times 10^7}{\eta_0 (DT)^2} c \phi(A' \sqrt{c}). \end{aligned} \quad (9)$$

In these equations, λ_1^0, λ_2^0 are the limiting equivalent conductances of the ions of the electrolyte, D the dielectric constant of the medium, η_0 the viscosity, and c the concentration of the electrolyte in mols per liter. When the concentration is zero, EQUATIONS 8 and 9 reduce to

$$\mathcal{D}_0 = 17.872 \times 10^{-10} T \left(\frac{\lambda_1^0 \lambda_2^0}{\Lambda^0} \right). \quad (10)$$

The term, in EQUATION 8, containing the activity coefficient on the mol per liter scale, may be computed by

$$\left(1 + c \frac{\partial \ln y_{\pm}}{\partial c} \right) = 1 - \frac{1.1514 \mathcal{S}_0 \sqrt{c}}{(1 + A' \sqrt{c})^2} + 4.606 Bc - c\psi(d) \quad (11)$$

\mathcal{S}_0 is the familiar Debye and Hückel limiting function and $A' \sqrt{c} = \kappa a$, where a is the mean distance of approach of the ions in centimeters and κ is the Debye and Hückel reciprocal distance. B is an empirical constant. The term, $c\psi(d)$ may be evaluated from

$$\psi(d) \equiv \frac{\partial d / \partial c + 0.001(2M_1 - M_2)}{d + 0.001c(2M_1 - M_2)} \quad (12)$$

where d is the density of the solution, M_1 the molecular weight of the solvent and M_2 that of the electrolyte.

* The numerical values in this and the following equations have been computed from the constants given by R. T. Birge (Rev. Modern Phys.: 13, 233. (1941)).

EQUATION 9 contains two concentration dependent terms which are the result of electrophoresis. The first of these contains the square of the difference in the limiting ionic conductances, and the second contains the function $\phi(A' \sqrt{c})$, which has the form of $c \log c$. Values of the function $\phi(A' \sqrt{c}) = \phi(\kappa a)$ are tabulated by Harned and Owen⁶ and also by Harned.⁷

For potassium chloride solutions, whose ions have nearly the same mobilities, the term containing the difference in ionic mobilities contributes only 0.0001 to the diffusion coefficient ($\mathfrak{D} \times 10^5$) at the highest concentration employed by us. It was for this reason that potassium chloride was selected, for, by this choice, not only was the situation simplified but the opportunity to investigate the $\phi(\kappa a)$ term directly was made possible.

Outline of the Experimental Method and Technical Precautions^{2, 3}

The method utilizes a cell in the form of an accurately machined rectangular parallelepiped provided with electrodes at the top and bottom in positions indicated by theory. The salt is introduced from a salt solution by a sliding mechanism at the bottom of the cell and diffuses upward either into water or a more dilute solution of the salt. The sliding mechanism minimizes turbulent flow. Electrodes on two sides of the cell are fixed at the same distance from both the top and bottom of the cell. The cell is completely filled with solution, so that the case is one of restricted diffusion.^{7, 9}

Introducing the equivalent concentration, c , into EQUATION 1 and combining with the equation of continuity

$$\frac{\partial c}{\partial t} = -\frac{\partial(cv)}{\partial x} \quad (13)$$

we obtain Fick's second law

$$\frac{\partial c}{\partial t} = \frac{\partial}{\partial x} \mathfrak{D} \frac{\partial c}{\partial x}. \quad (14)$$

Since we carry out the measurements as the diffusion proceeds and under conditions where the difference in concentrations at the top and bottom of the cell is small and decreasing with time, it is safe to regard the diffusion coefficient, \mathfrak{D} , as constant and employ Fick's law in the form

$$\frac{\partial c}{\partial t} = \mathfrak{D} \frac{\partial^2 c}{\partial x^2} \quad (15)$$

where c is the concentration at a vertical distance, x , with origin at the bottom of the cell. Let the depth of the cell be " a ". The flow is zero at $x = 0$ and $x = a$, and therefore, by EQUATION 1, we have the boundary condition

$$\frac{\partial c}{\partial x} = 0 \begin{cases} x = 0 \\ x = a. \end{cases} \quad (16)$$

With these restrictions, the general solution of EQUATION 15 is

$$c = \sum_{m=1}^{\infty} A_m e^{-m^2 \pi^2 \mathcal{D} t / a^2} \cos \frac{m \pi x}{a} + c_0 \quad (17)$$

where the A_m are Fourier coefficients. All even terms of this series vanish when the difference in concentrations at the bottom, $c(\xi)$, and the top, $c(a - \xi)$, is computed, and, if the electrodes are $\frac{1}{8}$ the depth of the cell from the top and bottom, the third term in the series vanishes. As a result, we may readily obtain the greatly simplified expression

$$c(\xi) - c(a - \xi) = 2A_1 e^{-\pi^2 \mathcal{D} t / a^2} + 2A_5 e^{-25\pi^2 \mathcal{D} t / a^2} \quad (18)$$

since ξ and $(a - \xi)$ are fixed and consequently the trigonometric functions are constant. Obviously, the second term on the right is negligible. For practical purposes, the logarithmic form

$$\ln (c(\xi) - c(a - \xi)) = -\frac{t}{\tau} + \text{constant} \quad (19)$$

is preferable where $a^2/\pi^2\mathcal{D}$ is replaced by τ .

For many electrolytes, the assumption can be made that, throughout the concentration range involved, the difference in concentrations $c(\xi) - c(a - \xi)$ is proportional to the difference in conductance ($K_B - K_T$). Indeed, Harned and French showed that, for potassium chloride at 0.002 to 0.01 molal, this assumption is valid to within 0.02%. EQUATION 19 now becomes

$$\ln [K_B - K_T] = -t/\tau + \text{constant} \quad (20)$$

whence the slope of the plot of the left of this equation *versus* t is $-1/\tau$ and

$$\mathcal{D} = \frac{a^2}{\pi^2} \frac{1}{\tau}. \quad (21)$$

The simplicity and directness of the method is apparent. All that is required are measurements of the conductance at the top and bottom of the cell at suitable time intervals, and a measurement of the depth of the cell. Of course, due to the inevitable imperfections in the mechanical construction of the cell, a slight correction is required for the ratio of the cell constants of the top and bottom pairs of electrodes.

The experimental technique has been described in detail elsewhere.^{2, 3} To assure the success of these experiments, great care was exercised to reduce the effect of vibration, to eliminate temperature gradients, and to guard against local heating at electrodes during the balancing of the conductance bridge.

After introduction of the salt, diffusion was allowed to proceed for 24 to 36 hours, and then five measurements of the conductances at the top and bottom of the cell were made each day. This procedure was continued for

five days. The solution was then stirred by placing a heating lamp near the cell. Enough radiation was absorbed by the platinum black on the electrodes to start convection currents and produce thorough mixing in a few hours. The final (infinity) readings were used to obtain the cell constant ratio correction in the manner described in the earlier communications.^{2,3} The method yields the differential diffusion coefficient at the concentration of the salt when t is infinity. To determine this concentration, the final solution was removed and its conductance measured in a small cell placed in an oil thermostat at 25°.

Experimental Results

The diffusion coefficient was computed directly from the experimental observations. The slope, $1/\tau$, was obtained from conductance readings at 24-hour intervals, and then the diffusion coefficient was computed by the equation

$$\mathcal{D} = \frac{a^2}{\pi^2} \left[\frac{\Delta \ln (K_B - K_T)}{\Delta t} \right]_{\Delta t = 86,400 \text{ secs.}} \quad (22)$$

Since five measurements of $(K_B - K_T)$ were made each day, five determinations of \mathcal{D} were obtained for the first to second, second to third etc. days, respectively. A typical result of a five-day experiment is illustrated by FIGURE 1, in which $\mathcal{D} \times 10^5$ is plotted against the number of an observation.

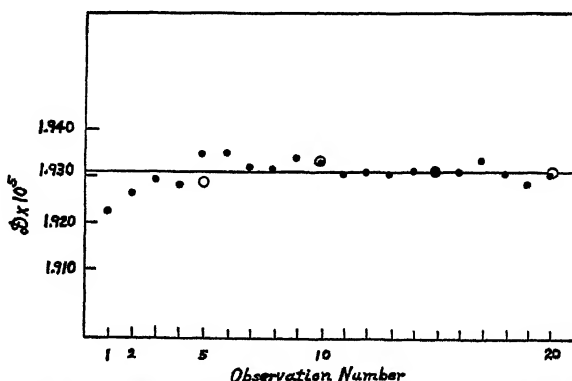


FIGURE 1. The diffusion coefficient of potassium chloride at 25° as derived from the slope,

$$\left[\frac{\Delta \ln (K_B - K_T)}{\Delta t} \right]_{\Delta t = 86,400 \text{ secs.}}$$

Dots represent the individual observations, the circles the mean of the preceding five observations, and the line the mean of all observations.

The inked circles represent the individual determinations and the circles the mean of the five determinations from results of the first to second day, the second to the third day, etc. The line represents the mean of all the determinations. The maximum deviation from this mean is 0.4 per cent and the

average deviation is 0.09 per cent, which is a very satisfactory result for a rate measurement.

The results at concentrations from 0.00125 to 0.5276 mols per liter are recorded in the third column of TABLE 1 and plotted against $c^{1/2}$ in FIGURE 1.

TABLE 1
OBSERVED AND CALCULATED DIFFUSION COEFFICIENT OF POTASSIUM CHLORIDE IN WATER AT 25°

Cell	C	$\mathcal{D} \times 10^5$ (obs)	$\mathcal{D} \times 10^5$ (theory)	Electrophoresis
	0.00	—	1.9958	—
A	.00125	1.961	1.960	0.002
A	.00194	1.954	1.953	.003
A	.00325	1.943	1.943	.005
A	.00585	1.931	1.929	.007
A	.00704	1.924	1.924	.008
A	.00980	1.918	1.915	.010
B	.01261	1.908	1.907	.011
C-5	.02654	1.879	1.883	.016
C-4	.03992	1.877	1.870	.020
C-3	.0462	1.872	1.866	.022
C-2	.0545	1.860	1.861	.024
C-1	.06074	1.856	1.858	.025
D	.1298	1.838	1.840	.034
D	.3323	1.842	1.839	.044
D	.5276	1.852	1.853	.048

Maximum effect of the term containing $(\lambda_1^0 - \lambda_2^0)$ in EQUATION 9 at $c = 0.5276$ is 0.0001. Last column contains values of the ϕ ($\lambda\phi$) term in EQUATION 9. Approximate areas of electrodes $A = 0.2 \times 2 \text{ cm}^2$; $B = 0.1 \times 2 \text{ cm}^2$; $C-5, C-4, C-3 = 0.2 \times 1 \text{ cm}^2$; $C-2, C-1 = 0.1 \times 1 \text{ cm}^2$; $D = (0.05)^2 \text{ cm}^2$.

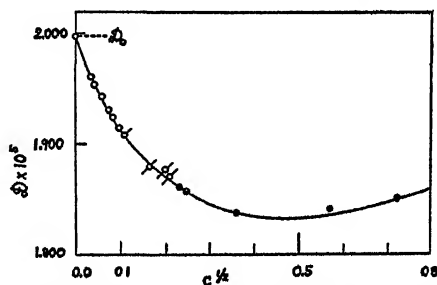


FIGURE 2 The diffusion coefficient of potassium chloride at 25° as a function of its concentration. The solid line was derived from theory.

The first six results were obtained in cell *A* with electrodes of areas of $0.2 \times 2 \text{ cm}^2$. Cell *B* contained electrodes with areas equal to $0.1 \times 2 \text{ cm}^2$, cells *C-5*, *C-4*, and *C-3* had electrodes with areas of $0.2 \times 1 \text{ cm}^2$, while cells *C-2* and *C-1* possessed electrodes of $0.1 \times 1 \text{ cm}^2$ areas. The last three results at the higher concentrations were obtained from a cell containing electrodes 1 mm. in diameter. No difficulty was encountered from polarization, and all these cells behaved in a satisfactory manner.

The fourth column of this table contains the values computed by the com-

plete theory as represented by EQUATIONS 8, 9, 10, 11, and 12 after introducing the values: $\lambda_1^0 = 73.52$, $\lambda_2^0 = 76.34^{10}$, $\eta_0 = 8.949 \times 10^{-8}$, $D = 78.54$, $\delta = 3.8$, $B = 0.0101$.¹¹ The values of $(\lambda a) = A' \sqrt{c}$ were obtained from the table given by Harned and Owen.¹² The agreement between the observed and the theoretical results is remarkably good. The last column of the table contains the contribution of electrophoresis as computed from the $\phi(A' \sqrt{c})$ or the $c \log c$ term in EQUATION 9. It is to be noted that, at all concentrations, this term is greater than the difference between the observed and calculated results, and that at the higher concentrations it is many times greater. This agreement appears to be direct evidence for the validity of the $c \log c$ term. However, we are somewhat surprised at the agreement at the higher concentrations where deviations due to changes in relative viscosity¹³ and dielectric constant should be expected to interfere. On the other hand, at concentrations below 0.1 *N* we are sure of the result and find in it the first objective proof of the validity of Nernst's limiting equation for 1-1 electrolytes.³

Summary

(1) The diffusion coefficient of potassium chloride from 0.00125 to 0.5276 *M* at 25°C. has been determined with an accuracy of the order of ± 0.1 per cent.

(2) The results are in excellent agreement with the theory of Onsager and Fuoss throughout this range of concentration. The agreement at the higher concentrations was unexpected and may be due to the fact that the relative viscosity of potassium chloride solutions changes little with the salt concentration.

(3) In another contribution from this laboratory, it will be shown that the unmodified theory will not account for the variation of the diffusion coefficient of calcium chloride in very dilute solutions. This disagreement parallels the behavior of the transference number of calcium chloride.

Bibliography

1. Ann. N. Y. Acad. Sci. 1945. 46: (5):209-349.
2. HARNED, H. S., & D. M. FRENCH. 1945. Ibid. 267.
3. HARNED, H. S., & R. L. NUTTALL. 1947. J. Am. Chem. Soc., 69: 736.
4. ONSAGER, L., & R. M. FUOSS. 1932. J. Phys. Chem. 36: 2689.
5. HARTLEY, G. S. 1931. Phil. Mag. 12: 473.
6. HARNED, H. S., & B. B. OWEN. 1943. *The Physical Chemistry of Electrolytic Solutions*. Reinhold. New York.
7. HARNED, H. S. 1947. Chem. Rev. 40: 461.
8. DEBYE P., & E. HÜCKEL. 1923. Phys. Z. 24: 305.
9. LONGSWORTH, L. G. 1945. Ann. N. Y. Acad. Sci. 46: 211.
10. SHEDLOVSKY, T. 1932. J. Am. Chem. Soc. 54: 1423.
11. Ref. 6: 381.
12. Ref. 6: 130.
13. GORDON, A. R. 1945. Ann. N. Y. Acad. Sci. 46: 285.

SOME INTERACTIONS IN SOLUTIONS OF ELECTROLYTES

BY C. A. KRAUS

Metcalf Research Laboratory, Brown University, Providence, Rhode Island

Introduction

The purpose of this paper is to present experimental material that has a bearing on our understanding of phenomena that have their origin in interactions between ions and ions and between ions and polar molecules in solutions of electrolytes in various solvents. Originally, it was not intended to touch upon long-range interaction between ions, but certain experimental material, which seems to have been overlooked up to now, calls for brief consideration.

The subjects which I shall discuss, aside from that of long-range interactions, are: (A) the interaction of free ions with polar molecules, those of the solvent as well as of other substances; (B) the effect of solvent and other dipolar molecules on the short range interaction of ions with one another; and, finally, (C) interactions of ion dipoles with one another in solvents of very low dielectric constant and up to high concentrations.

Long-Range Interactions

The presently accepted theory of strong electrolytes as developed by Debye and Hückel¹ and Onsager² has been amply verified for aqueous solution and has proved invaluable in elucidating the intricate phenomena that characterize these solutions. The extension of this theory by Bjerrum³ and Fuoss⁴ so as, also, to take account of short-range interactions between ions, has clarified our understanding of phenomena that appear in and characterize solutions of electrolytes in solvents of lower dielectric constant.

In water, it is possible to study a considerable number of phenomena, reversible as well as irreversible, by precision methods. In most non-aqueous solvents, we are largely dependent on the study of a single irreversible phenomenon, their electrical conductance, for our understanding of the nature of solutions of electrolytes in them. The conductance of solutions is the only property that can be measured quite generally with a high degree of precision in all solvents under a wide range of conditions. In what follows, we shall be largely dependent on conductance measurements for such experimental material as may throw light on the problem of solutions of electrolytes in non-aqueous solvents. It is unfortunate that much of the work that has been done in the past is lacking in precision and measurements have not yet been carried to concentrations sufficiently low to permit the application of limiting laws.

The theory of Onsager, according to which the equivalent conductance of

an electrolyte varies as a linear function of the square root of concentration, the slope of whose graph is a known function of temperature, dielectric constant, and viscosity of the medium and the charge on the ions, has been amply verified for aqueous solutions. In a few instances, it has been verified for solutions in solvents of somewhat lower dielectric constant, such as nitrobenzene and methanol. In solvents of dielectric constant lower than 30, verification becomes uncertain because of ion association.

There are few solvents that have a dielectric constant higher than that of water, and most of these, such as hydrogen fluoride and formamide, are not suitable for study because of instability or interaction with the dissolved substances. There is, however, one solvent which has a dielectric constant much higher than that of water, and which is stable, readily purified, and an excellent solvent for ordinary salts. I refer to liquid hydrogen cyanide, whose dielectric constant is 118.3 at 18° and 158.1 at 0°.⁵ Coates and Taylor⁶ measured the conductance of some 38 salts in this solvent at 18° more than a dozen years ago. The concentration range covered was from $2 \times 10^{-3}N$ to $1 \times 10^{-4}N$. Their results are remarkably consistent: the mean deviation of limiting conductances for several dozen salt pairs with a common ion is not above 0.5 per cent, and independent series of measurements are in good agreement.

The conductance of these solutions when plotted against the square root of concentration follows a linear relation. In most instances, the slope of these curves approximates that predicted on the basis of Onsager's theory. For a few salts, notably of lithium, the slope is greater than the theoretical; this may be accounted for by ion association. In some instances, however, particularly with salts of large ions, the observed slope is markedly below the theoretical, in some cases as much as 30 per cent. Examples are shown in FIGURES 1 and 2.

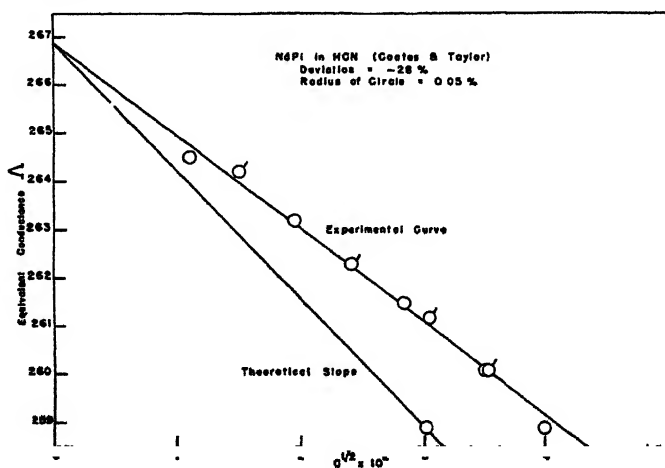


FIGURE 1. Plot of Λ against \sqrt{C} for sodium picrate in liquid hydrogen cyanide at 18°.

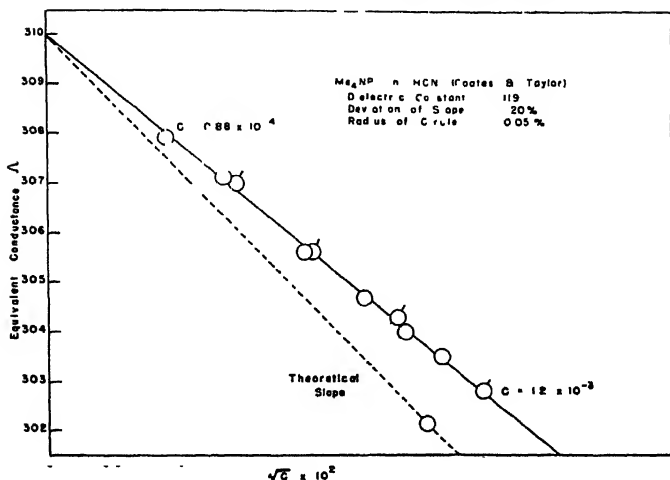


FIGURE 2. Plot of Λ against \sqrt{C} for tetramethylammonium picrate in hydrogen cyanide at 18°.

In FIGURE 1 is shown the graph for sodium picrate. The points lie on the straight line as drawn, with an average deviation of 0.05 per cent. The radius of the circles corresponds to an error of this value. The observed slope is 28 per cent below the theoretical. Results for two independent series of measurements are shown. In FIGURE 2, are shown results for tetramethylammonium picrate (2 series). The experimental values deviate from the straight line (in the mean) by 0.025 per cent. The slope deviates from the theoretical by -20 per cent.

It is difficult to understand how the experimental results could be in error by as much as 20-30 per cent. On the other hand, it is difficult to accept the conclusion that the underlying theory, which has been so adequately confirmed in aqueous solutions, should be in error. It should be borne in mind, however, that, aside from these measurements in liquid hydrogen cyanide, we have no information whatsoever concerning solutions in solvents of dielectric constant higher than that of water. Further investigation of solutions in liquid hydrogen cyanide is urgently called for.

Interaction of Ions with Solvent and Other Dipoles

Interactions with Solvent Molecules. The only means we have for studying the interaction of free ions with polar molecules is to measure their conductance. This means, in the first place, that precision measurements are carried to concentrations sufficiently low so that limiting conductances may be obtained by extrapolation according to established limiting laws. In the second place, since the quantity actually determined is the sum of the conductance of two ions, it is also necessary to know the transference numbers of the ions. These latter cannot be determined directly in most non-aqueous

solvents; accordingly, transference numbers must be evaluated by indirect means.

The basis of such evaluations is Stokes' law, according to which the speed of a particle moving under a given force through a given medium is inversely proportional to the viscosity of the medium. It has been recognized that Stokes' law can hold only for particles which are large with respect to the molecules of the medium through which they are moving. Walden⁷ has made use of this means for determining the conductance of ions in many different non-aqueous solvents. He has employed tetraethylammonium picrate for this purpose. Having determined the value of the conductance viscosity product for the picrate ion in water and knowing the viscosity of the non-aqueous solvent, the conductance of the picrate ion in that solvent becomes known.

In the Brown Laboratories, we have adopted a somewhat different, although related procedure.⁸ The conductance is measured of an electrolyte whose two ions are of the highest possible symmetry and which contain approximately the same (large) number of atoms other than hydrogen. It is then assumed that the two ions have the same conductance and, accordingly, that the conductance of each ion is one-half the total. As electrolyte, we have employed tetra-*n*-butylammonium triphenylborofluoride. Here, one ion contains 17 atoms other than hydrogen, the other, 20. It may be expected that the resistance experienced by the more closely packed atoms of the phenyl groups would be somewhat less than that of atoms in the chains of the butyl groups. The values obtained by means of the above-mentioned salt have been checked by measurements with octadecyltrimethylammonium octadecylsulfate.⁹ The results with the two salts agree within 5 per cent. It may be concluded that ion conductances based on the equal conductance of the tetrabutylammonium and the triphenylborofluoride ions are reliable within about 5 per cent.

In FIGURE 3 are shown values for the conductance viscosity product for the bromide and the sodium ions in seven and six different solvents, respectively. It is evident that the conductance of the bromide ion, relative to viscosity, varies widely. Most striking is the high conductance of this ion, as of most other simple ions, in water. The value of the $\Lambda_0^- \eta$ product for this ion in water is 2.7 times that in ethylene chloride. Even in water at 306°, the product is twice that in ethylene chloride at 25°. The exceptionally high conductance of simple ions in water is unique and has never been accounted for. Neither has anyone accounted for the fact that, with increasing temperature, the conductance of ions of high conductance increases less than that of ions of low conductance. The conductance of ions of very low conductance, such as the acetate ion, increases linearly with the fluidity of water.^{9a}

In FIGURE 3 are also shown values of the $\Lambda_0^+ \eta$ product for the sodium ion in six of the seven solvents listed in the figure. It will be noted that, in

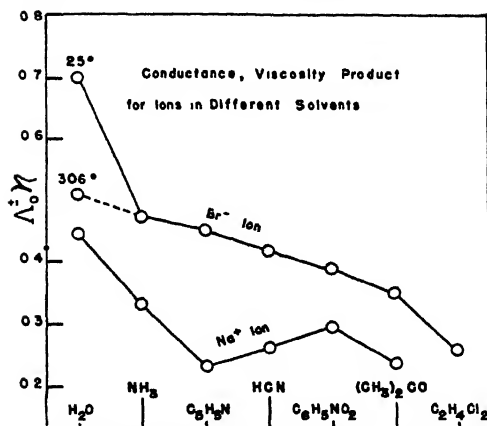


FIGURE 3. Conductance viscosity product for bromide and sodium ions in: water (306°)¹⁰, ammonia,¹¹ pyridine,¹² hydrogen cyanide,⁸ nitrobenzene,¹³ acetone,¹⁴ and ethylene chloride⁹

all cases, the conductance of the sodium ion is lower than that of the bromide ion—in some instances, much lower. It may be pointed out, here, that the conductance of negative ions is generally higher than that of similar positive ions.

The conductance of the sodium ion in water is much higher (relatively) than that in other solvents. The relatively low conductance of the sodium ion in pyridine is noteworthy.

In FIGURE 4 are compared $\Delta_0^+ \eta$ products for the tetraethylammonium

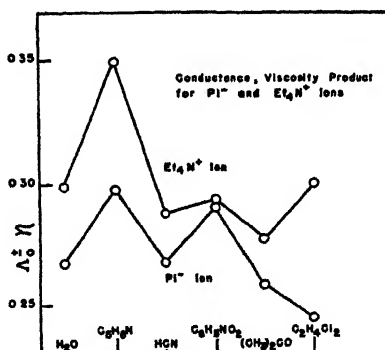


FIGURE 4. Conductance viscosity product, tetraethylammonium, and picrate ions in: water,¹⁵ pyridine,¹⁶ hydrogen cyanide,⁸ nitrobenzene,¹³ acetone,¹⁴ and ethylene chloride.⁹

and the picrate ions in six different solvents. It will be noted that the picrate ion is slower than the tetraethylammonium ion in all solvents, although the difference is small in nitrobenzene. The high conductance of both ions in pyridine and the wide spread of values in the case of pyridine and ethylene chloride is notable. In general, the lower conductance of the

picrate ion is understandable in view of the large number of atoms (other than hydrogen) that this ion contains. While the variation of the conductance viscosity product is smaller for these larger and more complex ions than it is for simple ions, it is, nevertheless, clear that the assumption that the $\Delta_0 \eta$ product is constant for the picrate ion is untenable.

It is of interest to compare the relative conductance of the three closely related halide ions: Cl^- , Br^- , and I^- . In FIGURE 5 are shown the differences

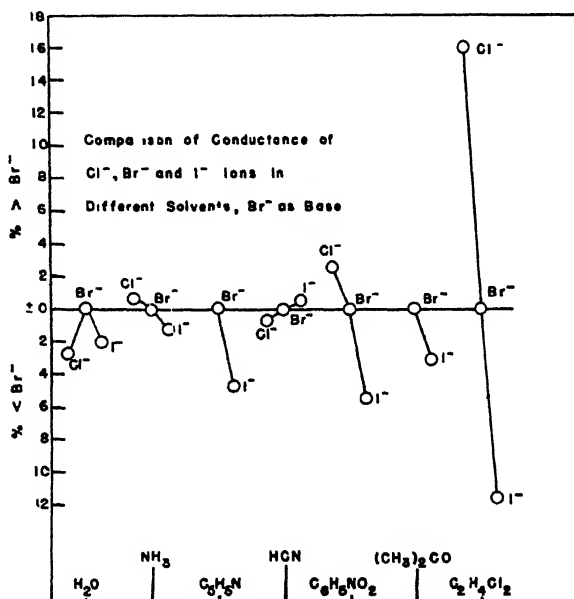


FIGURE 5 Comparison of Cl^- , Br^- , and I^- ion conductances in seven different solvents.¹⁶

(in per cent) between the Br^- and the Cl^- and I^- ions in seven different solvents.¹⁶ It will be noted that the order of conductance values for the three ions differs in different solvents. In water, the conductance of the Br^- ion is greater than that of either the Cl^- or the I^- ion, with that of the I^- ion slightly greater than that of the Cl^- ion. In ammonia, the conductance of the Cl^- ion is slightly greater than that of the Br^- ion which, in turn, is greater than that of the I^- ion. In hydrogen cyanide, the differences are exceptionally small, with the order reversed with respect to that in ammonia. But while the differences in conductance values are small for the three solvents just discussed, they are much larger for the four other solvents. The large spread in ethylene chloride is noteworthy: the conductance of the Cl^- ion is 16 per cent greater than that of the Br^- ion which, in turn, is 12 per cent greater than that of the I^- ion. The order of conductance values appears to be the same in pyridine, nitrobenzene, acetone, and ethylene chloride.

It is of interest to compare the conductance of small, composite ions of similar structure and comparable dimension. In FIGURE 6 are shown the

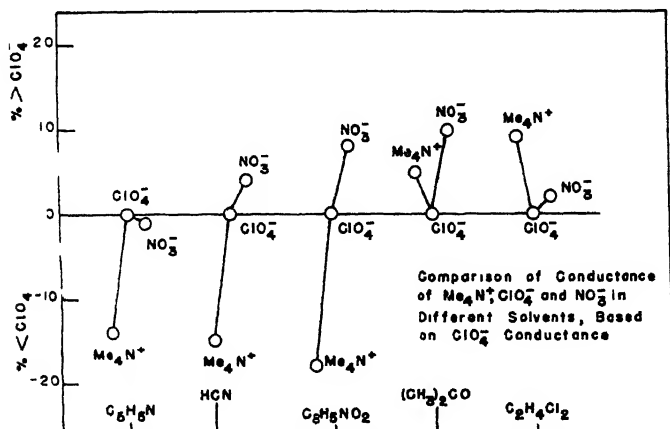


FIGURE 6 Comparison of $(\text{CH}_3)_4\text{N}^+$, ClO_4^- , and NO_3^- ion conductances in five different solvents.¹⁷

percentage differences between the conductance of NO_3^- , ClO_4^- , and $(\text{CH}_3)_4\text{N}^+$ ions, based on the conductance of the ClO_4^- ion, in five different solvents. The order of conductance values for these ions differs in different solvents.¹⁷ In pyridine, the conductance of the NO_3^- ion is slightly smaller than that of the ClO_4^- ion, while that of the Me_4N^+ ion is much smaller. In hydrogen cyanide and nitrobenzene, the conductance of the ClO_4^- ion is intermediate between that of the NO_3^- and the $(\text{CH}_3)_4\text{N}^+$ ions; in acetone and ethylene chloride, the conductance of the ClO_4^- ion is smaller than that of either of the other two ions. Differences are greatest in hydrogen cyanide and nitrobenzene, but they are marked in nearly all cases. It is of particular interest to note that, in hydrogen cyanide and nitrobenzene, the $(\text{CH}_3)_4\text{N}^+$ ion has a conductance nearly 20 per cent lower than that of the closely similar ClO_4^- ion, while in acetone and ethylene chloride, the conductance of the $(\text{CH}_3)_4\text{N}^+$ ion is up to 10 per cent greater than that of the ClO_4^- ion.

We must conclude that the free ions are highly individualistic in their interactions with solvent molecules. Aside from viscosity, there appears to be no correlation between these interactions and the physical properties of the solvents, such as dielectric constant, or with the size of the solvent molecules.

Effects Due to Ion Size and Structure. While the conductance of ions is dependent upon the number of atoms that they contain and upon the geometrical arrangement of these atoms within the ions, the manner in which the conductance of ions varies with their complexity and structure in different solvents is also dependent on the solvent medium. Studies of this kind are best carried out with the completely substituted onium ions, particularly

the quaternary ammonium ions. The conductance of the homologous series of quaternary ammonium ions from Me_4N^+ to $n\text{-Am}_4\text{N}^+$, inclusive, has been determined in ethylene chloride,¹⁸ nitrobenzene,¹⁹ and pyridine.²⁰ In addition, the conductance of the *n*-octadecyltributylammonium ion (30 C atoms) has been determined in the same three solvents; that of the dioctadecyldibutylammonium ion (44 C atoms) in ethylene chloride and nitrobenzene; and that of the dioctadecyldimethylammonium ion (38 C atoms) in nitrobenzene.²¹

In comparing the effect of size and structure on the conductance of ions as a function of their complexity, it is advantageous to compare not equivalent conductances but, rather, equivalent resistances, the reciprocal of ion conductances, as Thompson²² has suggested. Moreover, since the viscosity of solvents may differ greatly, comparison is simplified by multiplying the equivalent ion resistances by the reciprocal of solvent viscosity. In other words, we compare the reciprocals of the $\Delta_0^+\eta$ products.

In FIGURE 7, values of $1/\Delta_0^+\eta$ are shown as a function of the number of

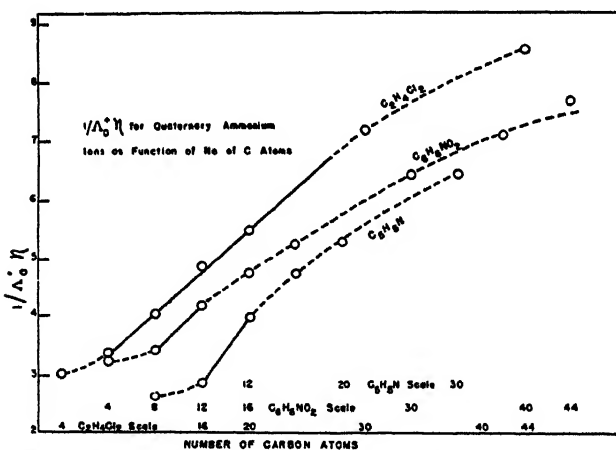


FIGURE 7. Reciprocal of conductance viscosity product for quaternary ammonium ions in ethylene chloride, nitrobenzene, and pyridine as function of number of carbon atoms.

carbon atoms in the cation for solutions in the solvents just mentioned. The number of carbon atoms in the cations is plotted along the axis of abscissas, the scale for each successive solvent being displaced four carbon atoms to the right to avoid confusion in the figure.

Perhaps the most striking feature of this graph is the fact that the equivalent resistance of quaternary ammonium ions in ethylene chloride increases as a linear function of the number of carbon atoms up to nearly 30 atoms. This means that every addition of CH_2 to the ion increases its resistance by the same amount, irrespective of chain length. The octadecyltributylammonium ion deviates very slightly toward lower resistance. The diocta-

decyldibutylammonium ion deviates more markedly, but the deviation is still surprisingly small considering the two long hydrocarbon chains and the dissymmetry of the ion.

For nitrobenzene, the curve between the ethyl and the propyl derivatives closely parallels that for ethylene chloride. Thereafter, it falls off and diverges widely. In the case of pyridine, the curve parallels that in ethylene chloride between the propyl and the butyl derivatives. It will be noted that the conductance viscosity product for the tetrabutylammonium ion has very nearly the same value in the three solvents. On the other hand, the resistance values of the methyl and ethyl derivatives are relatively much smaller in pyridine than in either of the other solvents. As the number of carbon atoms increases above 20, the curves for pyridine and nitrobenzene parallel each other rather closely.

It is evident that even large ions show specific differences and that Stokes' law does not apply. It may well be that the chains, so to speak, "curl up" so as to occupy a smaller effective volume in pyridine and nitrobenzene than in ethylene chloride. But we must be cautious in accepting any such simple hypothesis. Indications are, although data unfortunately are fragmentary, that ions with a large number of carbon atoms meet with greater resistance in water than in other solvents, even though ordinary ions meet with much smaller resistance. It is important to note the marked difference between the resistance of the tetramethyl and the tetraethyl substituted ions in pyridine and in ethylene chloride or nitrobenzene. It is difficult to escape the conclusion that Coulombic interaction with the solvent dipoles is greater in the latter two solvents than in pyridine.

Interaction with Dipoles Other than Those of the Solvent. It has previously been observed that the conductance of an electrolyte in a given solvent may be altered by addition of dipole molecules of some other substance in relatively small amount. Usually, this is due to a change in the dissociation of the ion pairs. However, in certain instances, the limiting conductance of an ion may be greatly altered by such additions. A study has been made of the effect, resulting from addition of ammonia, on the conductance of the Na^+ , Li^+ and Ag^+ ions in pyridine.²³ The conductance of all three ions is markedly increased on such addition. In FIGURE 8 is shown the result for the Li^+ ion on the addition of up to 0.37*N* of ammonia to a solution of lithium picrate in pyridine. The conductance of the Li^+ ion is increased by as much as 50 per cent. That the increased conductance is due to the lithium ion is shown by the fact that the conductance of tetrabutylammonium picrate is but little affected by the addition of ammonia¹² and that the conductance change is practically the same for sodium picrate and sodium iodide.²³ The conductance change for the Na^+ ion on addition of ammonia is nearly the same as that of the Li^+ ion; the conductance change of the Ag^+ ion is about one-half that of the Li^+ ion.^{23a} As may be seen from FIGURE 8, the addition

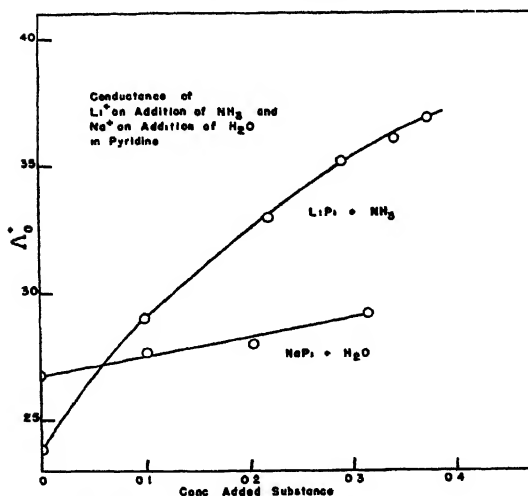


FIGURE 8. Effect of ammonia on the conductance of the lithium ion and of water on the conductance of the sodium ion in pyridine

of water to a solution of sodium picrate causes only a small change in the conductance of the Na^+ ion.

Since the conductance of an ion is a function of its dimensions and increases with decreasing dimension, the ammonia molecules must replace the larger pyridine molecules from association with the Li^+ or Na^+ ions to yield ions of smaller dimensions. It is surprising that the magnitude of the effect should be so large.

Effect of Interaction on Ion Association

In a given solvent, the association of ions to non-conducting ion-pairs is the smaller, the larger the ions.⁴ There is not, however, a clear correlation between ion association and ion conductance, except in the case of ions so large that their conductance is determined by the number of atoms permanently associated within the structure of the ions. The dissociation constant of quaternary ammonium salts increases with increasing number of carbon atoms in the substituent groups, but here, steric effects, also, must be taken into account. The dissociation constant of salts of quaternary ammonium ions does not increase greatly once the first two carbon atoms have been introduced into the substituent groups.

The factual material relating to specific effects due to structural features of ions is too copious and intricate to permit detailed presentation here. We shall content ourselves with a few cases in which interactions between ions and solvent molecules have a marked effect on the dissociation constant.

Effects Due to Interaction of Ions with Solvent Molecules. In TABLE 1 are given the dissociation constants of lithium, sodium, and potassium picrates

TABLE 1
DISSOCIATION CONSTANTS OF ALKALI METAL SALTS IN SEVERAL SOLVENTS

Salt	$K \times 10^4$		
	C_5H_5N	$C_6H_5NO_2$	$(CH_3)_2CO$
LiPi	0.806	0.0006	10.3
NaPi	.454	.28	13.5
KPi	1.054	.86	34.3
Bu ₄ NPi	12.8	2000	223.0

in pyridine,¹² nitrobenzene,¹³ and acetone.¹⁴ The constant is likewise given for tetrabutylammonium picrate in pyridine²⁰ and nitrobenzene,¹⁹ as well as acetone.¹⁴

As may be seen from TABLE 1, the alkali metal picrates are exceptionally weak electrolytes in nitrobenzene ($D = 34.5$). While a strong electrolyte like tetrabutylammonium picrate has a constant greater than 0.2, lithium picrate has a constant of only 6×10^{-8} . Sodium and potassium picrates are stronger than lithium picrate, but still very weak. For these two ions (Na^+ and K^+), there is probably some interaction between the solvent molecules and the ions, as they interact with the picrate ion at short range. The dissociation constant for the alkali metal picrates in pyridine ($D = 12.3$) is greater than in nitrobenzene. In pyridine, lithium picrate is stronger than sodium picrate, indicating greater interaction with the solvent molecules. However, all the alkali metal picrates are weak when compared with tetrabutylammonium picrate. In acetone ($D = 20.5$), sodium picrate is slightly stronger than lithium picrate and potassium picrate is much stronger than either of these two. The relatively high value for potassium picrate in all three solvents indicates that the lattice size of the potassium ion may be a factor in short-range interactions in these solvents.

Effect of Constitution of Solvent Molecules. It is of interest to point out that the properties of solutions of an electrolyte are dependent on the constitution of the solvent molecules, other controlling factors being the same. The one example of this kind for which data are available is that of ethylene dichloride and ethylidene dichloride. Although these two solvents differ greatly in boiling point and viscosity, they have almost identical dielectric constants, 10.23 and 10.2. In view of the value of their dielectric constants, we should expect salts in solution in these solvents to have the same dissociation constant. Values for the same salt in the two solvents are given in TABLE 2.

It will be noted that the dissociation constant of *o*-chlorophenyltrimethylammonium perchlorate in ethylidene chloride²⁴ is only one-tenth that in ethylene chloride. It appears that a similar relation holds in the case of tetra-alkylammonium picrates.²⁵ Ion conductances appear to decrease less

TABLE 2
DISSOCIATION CONSTANT OF SALT IN ISOMERIC SOLVENTS

Solvent	D. Const.	$K \times 10^6$
(O—ClC ₆ H ₄) ((CH ₃) ₃ NCIO ₄)		
Ethylene chloride	10.23	4.45
Ethylidene chloride	10.2	0.452

with increasing number of carbon atoms in ethylidene chloride than in ethylene chloride.

The $\Delta\eta$ product for *o*-chlorophenyltrimethylammonium perchlorate in the two solvents has the values 0.574 and 0.576, respectively.

Here, we have two isomeric solvents having the same dielectric constant, solutions in which exhibit widely differing properties. It seems reasonable to suppose that the differences are due to differences in the manner in which the molecules of these solvents interact with ions dissolved in them.

Effect Due to Negative Atoms in Cation. The introduction of negative atoms or groups into the organic substituents of the cation has a definite, although not a large, effect on the dissociation constant. This is illustrated by the values presented in TABLE 3, where constants are given for a number

TABLE 3
EFFECT ON DISSOCIATION CONSTANT OF INTRODUCING NEGATIVE ATOMS INTO CATION

Salt	$K \times 10^4$		
	C ₂ H ₄ Cl ₂	C ₆ H ₅ N	C ₆ H ₅ NO ₂
(C ₂ H ₅) (CH ₃) ₂ N Pi	0.460	8.21	440
(HOC ₂ H ₄) (CH ₃) ₂ N Pi	.066	9.50	70
(CH ₃ OCH ₂) (CH ₃) ₂ N Pi	.264	—	240
(BrCH ₂) (CH ₃) ₂ N Pi	.110	4.79	120
(CH ₃) ₂ N Pi	.320	6.71	400

of salts in ethylene chloride,²⁸ pyridine,¹² and nitrobenzene.¹³ Constants are also given for ethyltrimethylammonium and tetramethylammonium picrates for purposes of comparison.

The introduction of an OH group on the end carbon atom of the ethyl group of ethyltrimethylammonium picrate reduces the constant of the salt to about one-sixth its value in ethylene chloride and nitrobenzene. The introduction of a methoxymethyl group in place of the isomeric hydroxyethyl group yields a salt whose constant has about one-half the value of the ethyltrimethylammonium salt in ethylene chloride and nitrobenzene. Introducing a bromine atom into one of the methyl groups of the tetramethylammonium salt reduces the constant to about one-third its value in the same two solvents. But while the effects due to the introduction of these atoms and

groups closely parallel each other in ethylene chloride and nitrobenzene, the corresponding effects in pyridine are either very small or even reversed. Thus, the constant for hydroxyethyltrimethylammonium picrate is greater than that of the unsubstituted salt. Again, we find that the interactions between ions and solvent molecules are specific.

Effects Due to Hydrogen Bonding. In certain solvents, the interaction between ions due to Coulombic forces is markedly reinforced by hydrogen bonding of the negative ion if an active hydrogen atom is present in the cation. This is illustrated by the constants given in TABLE 4.¹³

TABLE 4
EFFECT OF HYDROGEN BONDING ON DISSOCIATION CONSTANTS

Salt	$K \times 10^4$	
	C_6H_5N	$C_6H_5NO_2$
Bu ₄ N Pi	12.5	>2000
Bu ₃ NH Pi	—	1.90
BuNH ₂ Pi	—	1.49
NH ₄ Pi	—	1.46
Ph·Me ₂ (OH)N Pi	12.8	0.20

The dissociation constant of tributylammonium picrate in nitrobenzene is less than one-thousandth that of the corresponding tetrabutylammonium salt (from 0.2 to 1.9×10^{-4}). Further replacement of butyl groups by hydrogen causes but little further change. The dissociation constant of phenyl-dimethylhydroxyammonium picrate in nitrobenzene is only 2×10^{-5} ; here, the bonding effect is very marked.

These salts which show bonding effects in nitrobenzene are such weak electrolytes in ethylene chloride that their constants cannot be evaluated with any considerable degree of certainty. In pyridine they are all normally strong electrolytes.¹² Thus, the hydroxy derivative has a constant of 12.8×10^{-4} , while the corresponding tetrabutyl derivative has a constant of 12.5×10^{-4} . The absence of bonding effects in pyridine is readily accounted for; pyridine being a basic solvent, its molecules bond to the hydrogen of the cation, thus precluding bonding to the negative ion.

Effects Due to Polar Molecules Other than Those of the Solvent. The addition of polar molecules to solutions of ordinary salts in a given medium may increase or decrease the dissociation constant or may leave the constant unchanged. The percentage change in the constant for a number of salts in pyridine on addition of ammonia is shown in column 3 of TABLE 5.^{26a}

The constant for lithium picrate is greatly increased on addition of ammonia even though the free lithium ion has effectively smaller dimensions as indicated by its greatly increased conductance. On the other hand, the constant for sodium picrate remains unchanged, although, as with the lithium ion, the conductance of the free ion is greatly increased. In the

TABLE 5
EFFECT OF AMMONIA ON DISSOCIATION CONSTANT OF SALTS IN PYRIDINE

Salt	Conc. NH_3	$10^2 \times \Delta K/K_0$	$10^2 \times \Delta \Lambda_0^+/\Lambda_0^+$
LiPi	0.22	+75	33
NaPi	.22	+0 0	46
NaI	.23	-27	44
AgNO ₃	.34	-56	19
Bu ₄ N Pi	.15	0.0	1.6*

* $10^2 \Delta \Lambda_0/\Lambda_0$.

case of sodium iodide and silver nitrate, the constant is greatly diminished, while the ion conductance is markedly increased for the sodium as well as the silver ion. The dissociation constant of tetrabutylammonium picrate remains unchanged on addition of ammonia. The total conductance of the salt is slightly increased, probably due to decreasing viscosity of the solvent medium.

The addition of 0.007 *M* of water increases the constant of sodium picrate 10% and leaves Λ_0 unchanged. The addition of 0.275 *M* methanol increases *K* 32 per cent and decreases Λ_0 but 1.17 per cent.¹²

We shall now consider the case of salts in which hydrogen bonding may occur. The hydrogen of the OH group in the trimethylhydroxyammonium ion is very active and, in the absence of basic molecules, as in nitrobenzene, it bonds with the negative ion to yield a very weak electrolyte. If basic molecules are added to such a solution, they interact with the active hydrogen of the ion and the solute behaves like a stronger electrolyte; the stronger the added base, the more marked is this effect. In FIGURE 9 are shown graphs

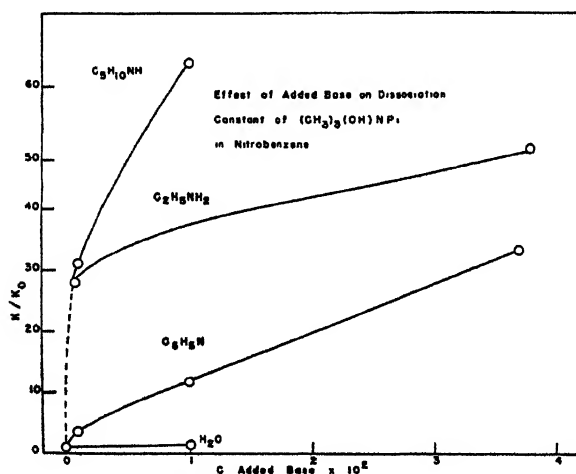


FIGURE 9. Effect of bases on the dissociation constant of trimethylhydroxyammonium picrate in nitrobenzene.

for the dissociation constant of trimethylhydroxyammonium picrate on addition of different bases;¹⁸ the dissociation constant of the salt in pure nitrobenzene is 0.17×10^{-1} . In FIGURE 9 are shown values of the ratio of the dissociation constant, K/K_0 , for this salt on addition of different bases.

In the case of water, a very weak base, the addition of 0.01 *M* causes an increase of only 30 per cent in the constant. A similar addition of pyridine increases the constant about ten times that of ethylamine, 35 times, and that of piperidine, 60 times. It is apparent that the increase of the dissociation constant of the solute is a function of the strength of the added base and increases greatly with increasing strength.

Interaction in Solvents of Very Low Dielectric Constant

In solvents of very low dielectric constant, solvent molecules have very small moments, while Coulombic forces are large. The properties of electrolyte solutions in solvents of this kind are influenced but little by the solvent medium: they are primarily determined by the size and structure of the ions. It is obvious that electrolytes, particularly those with very large ions, will be dissociated into their constituent ions to some (although small) extent at accessible concentrations. We should expect, however, that the properties of such solutions would be exceptionally sensitive to constitutional and dimensional factors.

The dissociation of a strong salt (with large ions) in benzene at 25° and at a concentration of $10^{-1}N$ is of the order of 4×10^{-1} per cent.²⁷ As a result of Coulombic interaction, such salts exist largely as ion-pairs at low concentrations.⁴ Values of polar moments of salts at low concentration bear this out.²⁸

At somewhat higher concentrations, ions of more complex type (triple ions) make their appearance, with the result that the equivalent conductance passes through a minimum.²⁹ At still higher concentrations, more complex aggregates are formed. The complexity of these aggregates is dependent on the dimensional and structural features of the ions themselves. In general, electrolytes with large ions exhibit a greater degree of dissociation into ions at very low concentration than do salts with small ions. They also exhibit a larger degree of association to more complex structures at higher concentrations.

Conductance as a Function of Concentration. In FIGURE 10 are shown conductance curves for tetra-*n*-butylammonium, tetra-*i*-amylammonium, and dioctadecyldi-*n*-butylammonium thiocyanate²⁷ in benzene at 25°. The logarithm of equivalent conductance is plotted as a function of the logarithm of concentration. The measurements were not carried to concentrations below that of minimum conductance. At sufficiently high concentrations, the conductance curves pass through a maximum due to the rapidly increasing viscosity of the solutions. Between $1N$ and $1 \times 10^{-4}N$, the curves exhibit very complex structures.

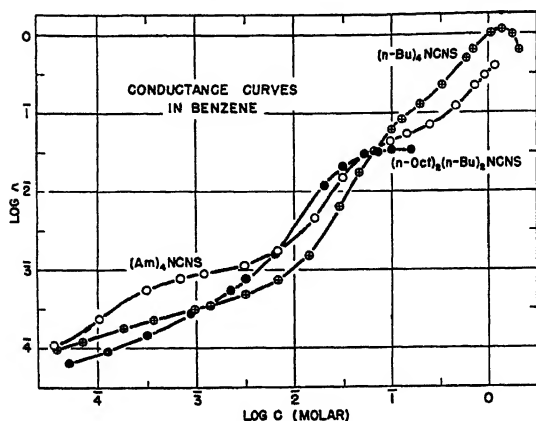


FIGURE 10. Conductance of some salts in benzene at 25°.

At high concentrations of salt, the concentration of free ions becomes appreciable. On the basis of the viscosity of benzene, we should expect the limiting equivalent conductance of electrolytes to be somewhat under 100. On this basis, the dissociation of tetrabutylammonium thiocyanate at 1*N* would be in the neighborhood of 1 per cent. But the viscosity of the 1*N* solution of salt is in the neighborhood of 8 or 9 times that of pure benzene. Therefore, the concentration of ions must be considerable. On the other hand, if multiple-charged aggregates are present, the conductance would be increased due to the greater transport value of the multiple-charged ions, much as in the case of colloidal electrolytes in water.

Molecular Weight and Association. Much light is thrown on the nature of these solutions by measurement of the molecular weight of the solutes by the freezing point method. The lowering of the freezing point of benzene due to a number of different salts, has been measured by a precision method.³⁰ In FIGURE 11 are shown values of the association number for several salts as

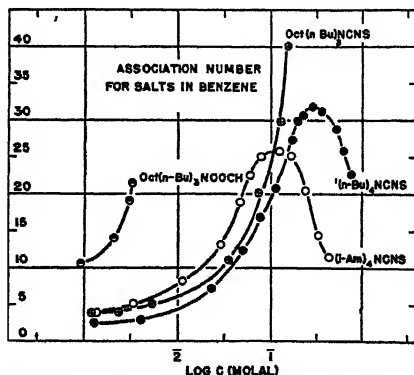


FIGURE 11. Association numbers of some salts in benzene at its freezing point.

a function of concentration.³¹ The association number is the average number of formula weights of solute per mole of solute as it exists in solution. Experimentally, it is the ratio of the freezing point lowering due to a normal solute, having the same formal concentration as the measured solute, divided by the freezing point lowering observed for the solute that has been measured.

As may be seen from FIGURE 11, all salts are associated at a concentration as low as 1×10^{-3} molal. In general, the smaller the substituent groups of the quaternary ammonium ion, the lower the degree of association. Thus, the association of tetraisoamylammonium thiocyanate at 10^{-3} molal is approximately twice that of the corresponding tetrabutylammonium salt.

As concentration increases, association increases and passes through a maximum. This maximum is displaced toward lower concentration as the number of carbon atoms in the substituent alkyl groups is increased. The introduction of long chains into the cation increases the degree of association markedly. This is shown in the case of both dioctadecyldibutylammonium formate and octadecyltributylammonium thiocyanate.³² With these salts, the measurements could not be carried to higher concentrations because the solubility limit was reached in each case.

The degree of association of salts in benzene is a function of the size and geometry of their ions. Salts with small ions exhibit much less association than do salts with large ions. The phenomena observed in solutions of salts in benzene are clearly due to interactions that have their seat in the charges on the ions. In this connection, it is of interest to note that the freezing point of a solution of octadecyl alcohol shows a deviation of only a few per cent from that of a normal solute.

Conclusion

The phenomena observed in benzene are not peculiar to this solvent. While there are no molecular weight data in solvents other than benzene, conductance data are available for concentrated solutions of salts in numerous solvents. Below a dielectric constant of 20, or somewhat lower, the conductance curves of all solutions exhibit a minimum and a steeply rising branch toward higher concentrations. In liquid ammonia ($D = 22$), some salts exhibit no minimum, others exhibit a minimum, and still others show little conductance change over a large concentration interval. In methylamine and ethylamine, all salts exhibit a minimum. If the concentration is carried high enough, the conductance passes through a maximum due to the increasing viscosity of the solution. It should be emphasized that the minimum and the increase of conductance toward higher concentrations is just as general and as characteristic for solutions in solvents of lower dielectric constant as is the lack of a minimum and the continuously decreasing conductance toward higher concentrations characteristic of solutions in solvents of higher dielectric constant, such as water.

Up to this time, the only theories we have that apply to solutions of electrolytes are of the limiting type. We have no theory that applies to concentrated solutions. The varied phenomena that may be observed in solutions of electrolytes have their origin in the charges on the ions. The observed phenomena are the resultant of interactions of the ions with one another and with molecules of the solvent medium or other molecules that may be present. It would seem that a solution of this problem of electrolytic solutions will be arrived at only through a microscopic analysis of the systems in question. In the meantime, there is need for much further experimental work of high precision with systems of the most varied nature, in order that the generality of various factors that underlie the observed phenomena may be determined.

Bibliography

1. DEBYE P., & E. HÜCKEL. 1923. Phys. Z. 24: 185.
2. ONSAGER, L. 1926. Phys. Z. 27: 388. 1928; 28: 277.
3. BJERRUM, N. 1926. Kgl. Danske Vidensk.; Selskab. 7: (9).
4. FUOSS, R. M., & C. A. KRAUS. 1933. J. Am. Chem. Soc. 55: 1019.
5. FREDENHAGEN, K., & J. DAHLMOSS. 1930. Z. anorg. Chem. 179: 77. COATES, G. E., & J. E. COATES. 1944. J. Chem. Soc.: 77.
6. COATES, G. E., & E. G. TAYLOR. 1936. J. Chem. Soc.: 1245, 1495.
7. WALDEN, P. & E. J. BIRR. 1931. Z. phys. Chem. 153A: 1.
8. FOWLER, D. L., & C. A. KRAUS. 1940. J. Am. Chem. Soc. 62: 2237.
9. THOMPSON, W. E., & C. A. KRAUS. 1947. J. Am. Chem. Soc. 69: 1016.
- 9a. KRAUS, C. A. 1922. *Electrically Conducting Systems*: 122. Chemical Catalog Co. New York.
10. NOYES, A. A. 1907. Carnegie Institution, Report 63.
11. HINZDA, V. F. 1935. Thesis, Brown University.
12. BURGESS, D. S., & C. A. KRAUS. 1943. J. Am. Chem. Soc. 70: 706.
13. WITSCHONKE, C. R., & C. A. KRAUS. 1947. J. Am. Chem. Soc. 69: 2472.
14. REYNOLDS, M. B. 1947. Thesis, Brown University.
15. EKWALL, P. 1933. Z. phys. Chem. 163A: 442.
16. References as for FIGURE 3.
17. References as for FIGURE 4.
18. TUCKER, L. M., & C. A. KRAUS. 1947. J. Am. Chem. Soc. 69: 1454.
19. TAYLOR, E. G., & C. A. KRAUS. 1947. J. Am. Chem. Soc. 69: 1731.
20. LUDIR, W. F., & C. A. KRAUS. 1947. J. Am. Chem. Soc. 69: 2481. CARIGNAN, C. J. 1947. Thesis, Brown University. BURGESS, Ref. 12.
21. PICKERING, H. L. 1947. Thesis, Brown University.
22. THOMPSON, W. E. 1941. Thesis, Brown University.
23. BURGESS, D. S., Ref. 12; CARIGNAN, C. J., Ref. 20.
- 23a. See, also, TABLE 5.
24. RAMSEY, J. B., & E. L. COLICHMAN. 1947. J. Am. Chem. Soc. 69: 3041.
25. MARTELL, A. E. Private Communication.
26. MEAD, D. J., J. B. RAMSEY, ROTHROCK, D. & C. A. KRAUS. 1947. J. Am. Chem. Soc. 69: 1528.
- 26a. CARIGNAN, C. J. Ref. 20; BURGESS, Ref. 12.
27. STRONG, L. E. 1940. Thesis, Brown University.
28. HOOPER, G. S., & C. A. KRAUS. 1934. J. Am. Chem. Soc. 56: 2265. GEDDES, J. A., & C. A. KRAUS. 1936. Trans. Faraday Soc. 32: 585.
29. FUOSS, R. M., & C. A. KRAUS. 1933. J. Am. Chem. Soc. 55: 2387.
30. KRAUS, C. A., & R. A. VINGEE. 1934. J. Am. Chem. Soc. 56: 511.
31. COPENHAFFER, D. T. 1942. Thesis, Brown University. Unpublished observations by Mr. H. S. Young in this laboratory.
32. YOUNG, H. S. Ref. 31.

DIELECTRIC ABSORPTION BY AMMONIUM SALTS IN SOLVENTS OF LOW DIELECTRIC CONSTANT

BY ROBERT H. COLE AND HOWARD A. STROBEL

Metcalf Research Laboratory, Brown University, Providence, Rhode Island

Introduction

The investigation of electrolyte behavior when the solvent has a low dielectric constant is of interest because electrostatic forces between ions become very large even at small concentrations. The importance of these forces has been made evident by a variety of measurements.¹

The existence of paired ions of unlike sign or of more complex polar groups at the expense of simple ions must lead to pronounced changes in the dielectric properties of such solutions. A simple ion-pair clearly constitutes a dipole with a moment depending on the size and symmetry of the ions concerned. It is, therefore, to be expected that the dielectric effects associated with polar molecules will have their counterpart in these solutions. This conclusion was confirmed, and with it the hypothesis of such ion dipoles, by dielectric constant measurements of Kraus *et al.*² for a number of ammonium salts. The enormously increased dielectric constants found beyond the critical concentrations required, on the assumption of a single species of dipole, dipole moments large compared even to strongly polar molecules.

Unfortunately for simplicity, more or less stable higher order groupings of ions may well exist, and a number of such species may exist simultaneously. It is important to realize that low frequency conductance and dielectric constant data, and cryoscopic values, can by their nature give only a statistical average of the effects of different species, presumably under near-equilibrium conditions. The dielectric absorption measurements to be discussed here* have an important advantage in this respect, in that they not only give a direct indication of the existence of ionic dipoles, but also show the speed with which they can react to an applied field. In this way, the complexity of the situation is made more directly apparent and indicates whether more specific calculations about the groups are possible and justified.

Theoretical Background

Absorption Formulae. The classic theory of dielectric absorption is that of Debye,³ in which molecular dipoles are pictured as being subject to the combined effect of an orienting torque resulting from any applied field and the electric fields of its neighbors. The mean equilibrium orientation under these influences is, however, built up only gradually as the molecules readjust

* The discussion is largely based on experimental work carried out as part of a general investigation of electrolytes in low dielectric constant solvents under C. A. Kraus. The absorption measurements were made by A. H. Sharbaugh, Jr., and H. A. Strobel. Part of the work reported was supported by a grant from the Research Corporation.

themselves to the altered situation. The effective field Debye approximates by the Lorentz field, and the readjustment by the model of a hydrodynamical sphere, subject to Brownian rotation, moving in the viscous fluid. With these assumptions, it is found that, for an applied sinusoidal field of frequency ν , the dielectric constant ϵ^* is given by

$$\epsilon^* = \epsilon' - i\epsilon'' = \epsilon_\infty + \frac{(\epsilon_0 - \epsilon_\infty)}{1 + \nu/\nu_c} \quad (1)$$

where

$$\nu_c = (\epsilon_\infty + 2)/2\pi\tau(\epsilon_0 + 2), \quad \tau = 6\eta V/kT. \quad (2)$$

In these equations, ν_c may be called the critical frequency of the absorption, expressed by the Debye model in terms of the molecular volume V and viscosity η , ϵ_0 is the static dielectric constant realized at low frequencies ($\nu \ll \nu_c$), and ϵ_∞ the dielectric constant at frequencies so high ($\nu \gg \nu_c$) that the dipole polarization does not form. The Boltzmann constant and absolute temperature are denoted by k and T .

The imaginary component ϵ'' gives rise to an "absorption" conductance as a result of dipole orientation which is unable to keep up with the alternations of the field, and the observed conductance is the sum of this and the ionic conductance. The relation between the extra absorption conductance $\Delta\kappa$ and ϵ'' is readily shown to be

$$\Delta\kappa = \nu\Delta\epsilon''/1.80 \times 10^{12}, \quad (3)$$

where $\Delta\epsilon''$ is used to indicate the value is for dipole absorption only. The measured conductance $\kappa(\nu)$ in a solution with ionic conductance κ_0 (assumed independent of frequency) is given by

$$\kappa(\nu) = \kappa_0 + \Delta\kappa = \kappa_0 + \frac{(\epsilon_0 - \epsilon_\infty)\nu/\nu_c}{1 + \nu^2/\nu_c^2} \frac{\nu}{1.80 \times 10^{12}}. \quad (4)$$

Although the critical frequency ν_c has been here presented as it appears from the Debye model by EQUATION 2, it should be noted that equations for ϵ^* and $\kappa(\nu)$, as given by EQUATIONS 1 and 4, result from a quite general assumption about dispersion processes,⁴ and different specific models lead merely to different significance of the parameter ν_c . Similarly, the quantity ϵ_0 is differently expressed in terms of molecular parameters, but equations of the same, or nearly the same, type form result. The dielectric constant and conductance EQUATIONS, 1 and 4, are thus quite generally representative of theoretical predictions. The frequency dependences of ϵ' , ϵ'' , and κ are shown by the solid curves in FIGURE 1, the scale of abscissae being logarithmic. We shall be primarily interested in the latter two, especially ϵ'' . It will be seen that the variation of ϵ'' on a logarithmic scale of frequency is symmetrical about a maximum at $\nu = \nu_c$, and that the conductance value

rises from the ionic value κ_0 to an inflection at $\nu = \nu_c$, reaching a final limiting value at frequencies high compared to ν_c .

Although the frequency variations shown in FIGURE 1 are quite a general

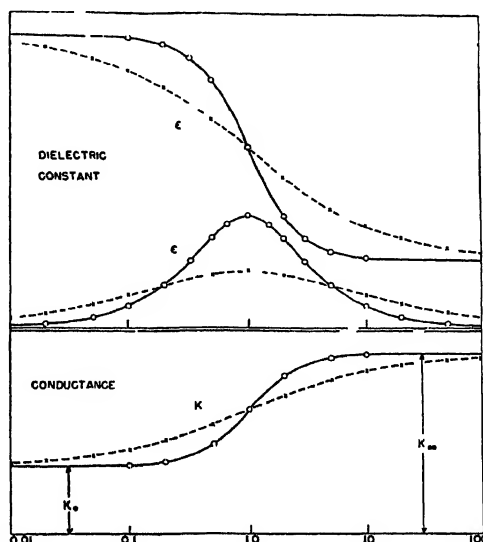


FIGURE 1. Dispersion and conductance curves plotted against logarithm of reduced frequency. Solid curves from Debye theory, dashed curves from EQUATION 5 for $\alpha = 0.4$.

prediction of simpler theories, it is important to recognize that experimental data on dielectrics often do not conform very well to the predictions. For many dielectrics, it is found that the actual curves retain a logarithmic symmetry but are spread out over a much wider band of frequencies, as indicated by the dotted lines of FIGURE 1. When this situation is found, any attempt to calculate molecular volumes by EQUATIONS 2 is clearly of dubious significance, as are attempts to calculate dipole moments from absorption data. (see the Discussion of "*Effects of Negative Ion Substitutions*"). Broader absorption regions encountered in polymer solutions have been quantitatively treated by Kirkwood and Fuoss⁵ on the basis of internal coiling of polymer chains, and the effects of variation in size of the polar unit leading to a set of discrete relaxation times have been discussed by a number of writers (see reference 6). These hypotheses are, however, not easy to reconcile with both the observed behavior and reasonable models in many cases. An as yet purely empirical representation for a large variety of dielectrics is found to be the addition of a parameter α ⁶ modifying the frequency dependent term in EQUATION 1, giving

$$\epsilon^* = \epsilon_\infty + \frac{\epsilon_0 - \epsilon_\infty}{1 + (i\nu/\nu_c)^{1-\alpha}} \quad (5)$$

The limits on α in this equation are zero, which gives EQUATION 1 as a limiting case, and unity, which gives a frequency spread of infinite width. The dashed curves of FIGURE 1 are drawn for $\alpha = 0.4$.

If the available frequencies ν are much larger or smaller than ν_c , only one end of the absorption region is accessible and only the limiting frequency behavior can be observed. These are easily shown from EQUATION 5 to be a simple power law dependence of $\Delta\epsilon''$ on ν :

$$\begin{aligned} \text{For } \nu < \nu_c, \Delta\epsilon'' &= (\epsilon_0 - \epsilon_\infty)(\nu/\nu_c)^{1-\alpha} \sin(1-\alpha)\pi/2, \\ &= (\epsilon_0 - \epsilon_\infty)(\nu/\nu_c) \text{ if } \alpha = 0. \\ \text{For } \nu > \nu_c, \Delta\epsilon'' &= (\epsilon_0 - \epsilon_\infty)(\nu_c/\nu)^{1-\alpha} \sin(1-\alpha)\pi/2, \\ &= (\epsilon_0 - \epsilon_\infty)(\nu_c/\nu) \text{ if } \alpha = 0. \end{aligned} \quad (6)$$

The approach of $\Delta\epsilon''$ to these limiting relations is conveniently shown on a double logarithmic plot against frequency, as is done in FIGURE 2. For $\alpha =$

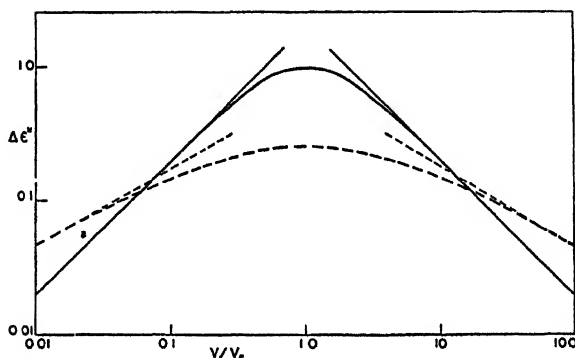


FIGURE 2. Logarithmic plot of absorption $\Delta\epsilon''$ vs. frequency. Solid curve from Debye theory, dashed curve from EQUATION 5 for $\alpha = 0.4$, straight line asymptotes from EQUATION 6.

0 (Debye type behavior), the absorption $\Delta\epsilon''$ varies either directly as, or inversely as, the frequency, except within roughly a factor of five of the critical value ν_c . Correspondingly, the absorption conductance from EQUATION 4 varies as the square of the frequency below ν_c , approaching constancy above ν_c . For the broader absorption curve (dashed line, $\alpha = 0.4$), the limiting behavior of $\Delta\epsilon''$ as a fractional direct or inverse power of frequency is less rapidly approached.

Calculation of Dipole Moments. An examination of absorption curves as plotted in FIGURE 2 can indicate, even though the complete curve is not obtained, whether the Debye type curve can represent the data, or if the curves predicted by EQUATION 5 can, by suitable choice of α , be made to fit, as they do in many dielectric problems.⁶ Once this is done, the value of $\Delta\epsilon''$ determines the value of $(\epsilon_0 - \epsilon_\infty)/\nu_c$ if $\nu \ll \nu_c$ or $(\epsilon_0 - \epsilon_\infty)\nu_c$ if $\nu \gg$

ν_c , and a knowledge of $\epsilon_0 - \epsilon_\infty$ from other data then determines ν_c . If, further, the curve is of Debye type form, EQUATIONS 2 then determine the Debye relaxation time τ and permit at least qualitative estimates of molecular volume V . If, however, the dispersion is not of this form, the application of EQUATION 2 to calculate a value of V is of very uncertain significance, and, failing an adequate dispersion theory, the quantity ν_c had perhaps best be left as a measured but unexplained parameter.

A further use of the absorption data is in calculation of dipole moments, as originally developed by Debye⁷ for the case of polar molecules in non-polar solvents. Generalization of the polarization formulae as a function of concentration leads to expressions for both ϵ' and $\Delta\epsilon''$ of the solution in terms of dipole moment and concentration of solute dipoles. Assuming that $\nu \ll \nu_c$ and that the solution is dilute enough that $\Delta\epsilon'' \ll \epsilon_0$, Debye's result⁷ is obtained that

$$\mu^2 = \left(\frac{3}{\epsilon_0 + 2} \right)^2 \frac{3kT}{4\pi N} \frac{M_1}{x_2 d_1} \frac{\nu_c}{\nu} \Delta\epsilon'' \quad (7)$$

where μ is the dipole moment in c.s.u.; M_1 and d_1 are molecular weight and density of the solvent; x_2 is the mole fraction of solute; and N is Avogadro's number.

The derivation of EQUATION 7 involves a number of assumptions which may be open to question in the present application, as follows:

- (1) The solute exists in the solutions only as dipoles of moment μ .
- (2) The absorption has the simple Debye-type frequency dependence and the sphere viscosity model is adequate.
- (3) The mechanism of solute polarization is adequately represented by the Lorentz local field.
- (4) The constants ϵ_0 and ϵ_∞ , as they appear in EQUATION 7 for μ^2 and in EQUATIONS 6 and 2 to determine τ from ν_c , are properly determined.

From this list of assumptions, it should be recognized that only in simple situations can the use of relations such as EQUATION 7 to determine dipole moments be quantitatively very exact. An important advantage of absorption results, in principle, is the fact that one can decide from data at a single concentration whether assumptions (1) and (2) can be valid.

The Calorimetric Method

The calorimetric method of determining dielectric losses in liquids is not new, having been first used by Harms in 1901.⁸ Since then, it has been applied by a number of other workers, and the application to polar molecules was treated by Debye.⁷ Most of this work has been done in investigation of polar liquids, and the first extensive application to electrolytes was made in a series of investigations at Brown. A detailed discussion of the method having been given elsewhere,⁹ only a cursory description is necessary here.

The basic principle of the method is very simple. A cell in the form of a thermometer with two cylindrical electrodes in the bulb is filled with the solution to a convenient height in the stem. When a known voltage is applied to the cell, heating due to dielectric losses and electrolytic conductance causes the meniscus to rise, the rate increasing as the square of the field and the total conductance. A direct calculation of the conductance in terms of heat capacities and thermal expansion together with correction for thermal losses is neither convenient nor precise. Whenever feasible, the cell is calibrated, instead, for each solution by measuring the rate of heat generation at an audio frequency for which the conductance is also determined by standard bridge methods. The calculation of conductance in the frequency range of absorption is thus referred to a low-frequency absolute value, which, furthermore, in many cases can be taken to be the ionic conductance κ_0 . The value of $\Delta\epsilon''$ is then given by EQUATION 3. Although radiation and conduction losses of heat and dielectric losses in the body of the cell are not wholly compensated for without correction, their effect is small in many cases and can be satisfactorily estimated by measurements employing pure solvents.

The outstanding advantages of the method are its simplicity and high sensitivity. The former permits measurements at high radio frequencies, and the second makes possible measurements of effects at low concentrations. There are, of course, limits of several kinds on the range of usefulness. At low frequencies, ionic conductance usually becomes large compared to the absorption conductance, and values for the latter are less inaccurate owing to their being obtained as a small difference.

Residual effects in the apparatus also set limits. If the absorption conductance is too small, cell dielectric losses make necessary large and necessarily approximate corrections, particularly at frequencies above 10 m.c. for glass cells. Another high frequency difficulty is that of correct measurement of voltage across the cell, owing to residual circuit inductance. These limitations are mild enough that the method has proved a versatile and useful tool capable of further extension. Measurements with the apparatus at Brown have proved satisfactory over the frequency range 100 kc. to 30 mc. for many systems. The data considered in what follows were obtained by Sharbaugh et al.¹⁰ and by Strobel.¹¹ All measurements were at 25°C.

Discussion of Experimental Results

Salts with Large Asymmetrical Ions. As outlined in the Introduction, the calorimetric method has been applied primarily to investigation of polar complexes formed by ammonium salts in low dielectric constant solvents, usually benzene. As might be expected, the simplest behavior has been found for salts with large and unsymmetrical ions. As examples, there are shown in FIGURE 3 the dielectric losses for three slightly soluble salts in which

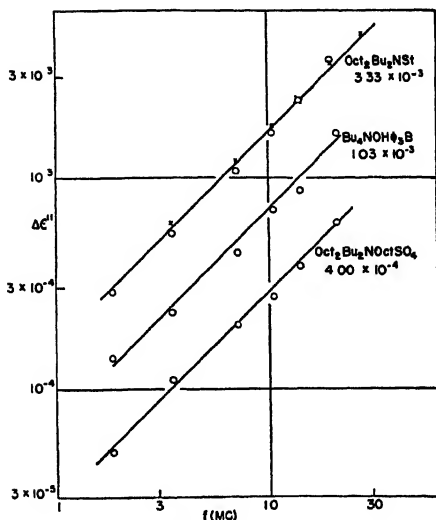


FIGURE 3. Logarithmic plot of absorption vs. frequency for large asymmetrical ion pairs in benzene.

the ions contain either long chain groups or are otherwise of low electrical symmetry. The solubility limited concentrations are, in all cases, somewhat greater than the critical "break" point from conductance data. The straight lines are drawn with a slope of unity on the double logarithmic plot, and the fit indicates that $\Delta\epsilon''$ varies linearly with frequency. This is in agreement with the Debye absorption theory prediction for frequencies much less than the critical frequency and makes reasonable its use in at least a semi-quantitative way to compute something about the nature of the polar aggregates.

It is natural to assume that only one species of aggregate exists, which, moreover, involves virtually the entire salt content. This is not unreasonable in view of the simple character of the loss-frequency curve and of conductance data indicating few ions at these concentrations, and is supported *a posteriori* by the results of calculations based on the assumption. The magnitude of the absorption at a given frequency then gives a value of $(\epsilon_0 - \epsilon_\infty)/\nu_c$ from EQUATION 6. The values of ϵ_0 and ϵ_∞ are determinable (with some reservation) by using audio-frequency values for ϵ_0 and the dielectric constant of the solvent for ϵ_∞ . The critical frequencies ν_c are then known, and, by inserting values of solution viscosity in EQUATION 2, values of dipole volume can be obtained. Similarly, the use of these data in EQUATION 7 yields a value of dipole moment.

The results for the three salts are shown in TABLE 1. It is seen that the calculated molecular volumes are of reasonable magnitude for ion pairs, though somewhat small, and that the very large dipole moments are compatible with pairing of large asymmetric ions.

TABLE 1
DIPOLE MOMENTS AND VOLUMES FROM ABSORPTION DATA

Salt, Concentration		$V(\text{\AA}^3)$	$a(\text{\AA})$	$\mu(D)$
$\text{Bu}_3\text{Oct}_2\text{NOctSO}_4$	(00040)	810	5.8	14.7
$\text{Bu}_4\text{NPh}_2\text{BOH}$	(0010)	990	6.2	13.5
$\text{Bu}_2\text{Oct}_2\text{NSt}$	(0033)	1900	7.3	9.0

Effect of Solvent and Concentration in Monodisperse Systems. The assumptions involved in obtaining values for V and μ are seen to be several. In particular, the results are based on the validity of the viscosity model for the relaxation process and the Lorentz field for the polarization. There are partial tests as to the consistency of these models; one based on varying concentration and solvent viscosity, and the other on agreement with molar polarization values from audio-frequency dielectric constant data. These tests have been made for tri-*iso*amylammonium picrate,¹¹ a salt which has large, electrically asymmetrical ions and is a weak electrolyte in benzene solution. The dielectric absorption results are shown in FIGURE 4a. The

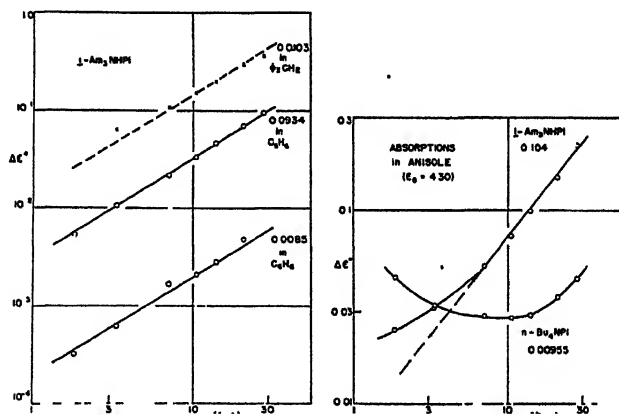


FIGURE 4a. (left) Variation of absorption for tri-*iso*-amylammonium picrate with concentration and solvent

FIGURE 4b. (right) Comparison of $\Delta\epsilon''$ for ternary and quaternary picrates in anisole.

solid lines on the plot of $\Delta\epsilon''$ are for two concentrations in benzene, and the dashed lines for a solution in diphenylmethane, a weakly polar solvent with nearly five times the viscosity of benzene.

It is qualitatively apparent from FIGURE 4a that the shifts of the curves are consistent with concentration increase causing a proportionate increase in $\Delta\epsilon''$, and the greater viscosity of diphenylmethane causing an increase by lowering the critical frequency ν_c and displacing the curve to the left. These conclusions are made quantitative by calculations of V and μ , a procedure

again made plausible by the linear variation of $\Delta\epsilon''$ with ν . The results, summarized in TABLE 2, show that variations of concentration by a factor of

TABLE 2
DIPOLE MOMENTS AND VOLUMES FOR TRI-*iso*-AMYLAMMONIUM PICRATE

Solvent, Concentration	$V(\text{\AA}^3)$	$a(\text{\AA})$	$\mu(D)$
C_6H_6 (.0085) (.093)	590	5.2	10.5
	460	4.8	10.4
CH_2Ph_2 (.010) (.069)	350	4.4	11.0
	860	5.9	9.7

ten and solvent viscosity by a factor of five yield reasonably consistent calculated values of dipole moment and volume. This consistency lends considerable support to the view that, for this salt, a single, well-defined dipole ion pair comprises virtually all of the salt content regardless of the large changes in concentration and viscosity. Such a unit is, of course, reasonable because of the asymmetry of the ions. This view is further in harmony with low-frequency dielectric constant data which yield a dipole moment of 13 debye units from the Debye polarization at infinite dilution, as compared to the absorption values of 9.7–11 debye units. Not only is the agreement fairly good; the difference is consistent with observed concentration effect on polarization as well.

That the situation is often less simple is illustrated by data on tetrabutylammonium picrate, differing from the previous case by having a symmetrical cation. In this case, anisole was used as the solvent, giving a strong electrolyte. Data for the tri-*iso*-amylammonium picrate in anisole are plotted, together with the tetrabutylammonium picrate data, in FIGURE 4b, to show the difference. It is seen that the former gives a linear $\Delta\epsilon''$ vs. ν relation as in the other solvents, except for a slight curvature at the lowest frequencies, which is probably experimental inaccuracy resulting from the large ionic conductance correction. The more symmetrical cation, however, results in a pronounced minimum. The values of $\Delta\epsilon''$ in this case are large enough, compared to the conductance correction, that there is little doubt of the effect being real.

On the basis of the Debye theory, a minimum of $\Delta\epsilon''$ can come about only by the overlapping of more than one dispersion region, the experimental data falling in the transition range between two such regions. Without a much more extensive frequency range, absorption data, clearly, cannot tell how much of the total dipole effect belongs to either region, to what extent the regions overlap, or even whether either region can be described by relations of the form of EQUATION 1. Under these circumstances, there is, clearly, no way to determine dipole moments or dipole volumes, and further measurements are necessary to establish the true situation and contributing

factors to it. One important point is illustrated by this example, namely, that absorption or other data involving the effect of time rates of change can show very directly the complexity of the dipole formation in a way that equilibrium measurements, representing a sum of effects, cannot, except by examining consistency with specific assumptions.

Picrate Data. The isolated example of tetrabutylammonium picrate clearly does not give any very complete indication as to the contributing factors in causing more complex dipole formation. A further attack on the question has been made by substitution of long chains for butyl groups in the ammonium picrate.¹¹ The data for octadecyl and dioctadecyl salts at several concentrations in benzene are plotted in FIGURE 5. In the graph

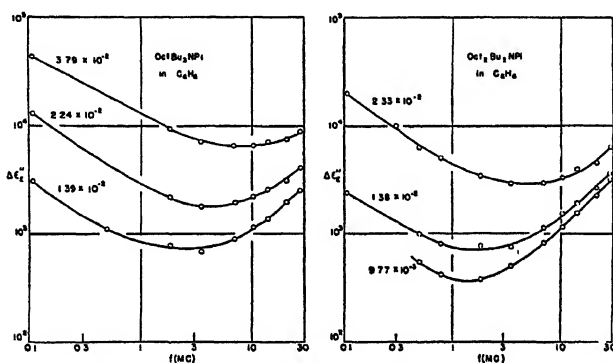


FIGURE 5. Equivalent absorption vs. frequency of substituted quaternary ammonium picrates at three concentrations in benzene.

are plotted values of equivalent absorption $\Delta\epsilon_B''$, defined as the specific absorption $\Delta\epsilon''$ multiplied by $1,000/c$, where c is the molar concentration. If the dipole absorption represents all or a constant fraction of the salt content, $\Delta\epsilon_B''$ should be sensibly independent of concentration. It is seen, from FIGURE 5, that this is by no means the case for either salt, $\Delta\epsilon_B''$ increasing markedly with concentration. This behavior is to be contrasted with that for tri-*iso*amylammonium picrate, where for a given solvent $\Delta\epsilon_B''$ is very nearly constant ($\Delta\epsilon''$ proportional to c).

Comparison of the long-chain substituted salts with the data of FIGURE 4 for tetrabutylammonium picrate shows that the $\Delta\epsilon_B''$ curves all show the general characteristic of a minimum between two absorption regions, which is modified only in detail by the concentration. The shift of the minimum to higher frequencies with larger salt concentration must be a result of shifts in position of the two absorption regions, change in their relative importance, or both. That the total contribution to dipole polarization increases more rapidly than salt concentration is plausible from the curves. This is confirmed by low-frequency dielectric constant values, which show a similar rapid increase above concentrations of 10^{-2} , a behavior not found for tri-*iso*amylammonium picrate.

On the basis of the Debye model, one would naturally associate the lower frequency absorption region with existence of very much larger dipole aggregates, and the curves of FIGURE 5 then suggest an increasing proportion of these as concentration increases. Without more complete data, this can, however, be no more than a reasonable guess. The simplest data to help clarify matters would undoubtedly be values for the dielectric constant ϵ' in the region of the minimum, which would give a measure of the net polarization at these frequencies. By comparison with low and high frequency values, limits could then be set on the contributions to the two types of polarization and areas under their absorption curves. A more complete frequency range of values for $\Delta\epsilon''$ would serve the same purpose, but the necessary instrumentation to accomplish this is not a minor undertaking.

Another factor, not as adequately investigated, is that of effects at very low frequencies. The quaternary picrate solutions all showed large apparent increases in ϵ' at audio frequencies (see the section on "Related Effects"). Whether this is real or the result of electrode effects is unknown; but if it is real, the situation becomes even more complex. It is to be noted that, in any such case, intercomparison with conductance measurements and freezing point or solvent distribution coefficient data must be made with care to insure one is talking about the same thing in all cases.

Effects of Negative Ion Substitutions. A partial indication of differences determined, to a great extent at least, by the negative ion in ammonium salts is evident in the results already discussed. For dioctadecyldibutyl salts, the octadecyl sulfate and stearate ions, both containing long chains, give every indication of a single, simple absorption region, and one is naturally led to the idea of a single homogeneous species of dipole, presumably an ion pair with large dipole moment. For the picrate salt, however, one must assume at least two reasonably well-defined species, which are otherwise somewhat indeterminate from present data. An extension of the argument would lead one to expect even more complex behavior and more extensive aggregation for smaller and more symmetrical anions, and this is indeed found.

The effect of anion substitution for strong electrolytes in benzene is shown by measurements in a number of octadecyltri-*n*-butylammonium salts, these being sufficiently soluble to permit significant variations in concentration. Substitution of iodide, chloride, nitrate, formate, and thiocyanate ions has been investigated¹¹ and, in all cases, a broad absorption region was found except at low concentrations for the frequency range of measurement. Examples of the results are shown in FIGURE 6 for the iodide and nitrate ions.

For the iodide at the lowest concentration ($c = 1.60 \times 10^{-3}$), it is seen that a straight line of unity slope fits the data reasonably well, which is consistent with the assumption that a single species of dipole exists with a high relaxation frequency. Calculations for this case based on EQUATIONS

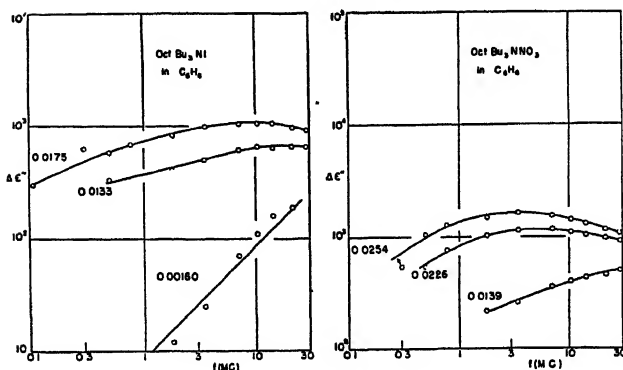


FIGURE 6. Equivalent absorptions vs frequency for iodide and nitrate substitutions as anions.

6 and 7 give values consistent with the hypothesis of an ion pair with large dipole moment. At higher concentrations, however, the situation is very different. The frequency ν_c is of the order 10–30 megacycles, instead of the calculated value of 1700 m.c. for the lowest concentration, and the curves are much too broad to be fitted by a Debye type absorption. The curves can be fitted quite well by the generalized empirical formula, EQUATION 5, using values of α from 0.40–0.60.

The values of α required to fit the shape of the curve can be compared with an alternative calculation based on the fact that the dispersion of dielectric constant and loss factor are not wholly independent. If EQUATION 5 is valid, it is readily shown that the maximum value $\Delta\epsilon''(\nu_c)$ of the absorption is given in terms of the dielectric constants ϵ_0 and ϵ_∞ as

$$\Delta\epsilon''(\nu_c) = (1/2)(\epsilon_0 - \epsilon_\infty) \tan(1 - \alpha)\pi/4. \quad (8)$$

From this, the value of α to give $\Delta\epsilon''(\nu_c)$ correctly may be found from appropriate values of ϵ_0 and ϵ_∞ . The curve for $c = 1.75 \times 10^{-2}$, for example, gives the value $\Delta\epsilon''(\nu_c) = 0.019$. Taking ϵ_∞ to be 2.272, the dielectric constant of the solvent, and ϵ_0 to be the dielectric constant at 7.5 kc gives $\alpha = 0.56$, as compared to a value $\alpha = 0.5$ obtained by best fit of the loss curve. Similar agreements are obtained for other concentrations. When it is realized that some uncertainty is attached to the propriety of the ϵ_0 and ϵ_∞ values used, one can conclude that the empirical fit is an adequate representation of the absorption data.

The very great difference in form of the curves of FIGURE 6 from the Debye result is evident from FIGURE 1, in which the curve for $\alpha = 0.40$ is compared with the limiting case $\alpha = 0$. Under these circumstances, the use of EQUATIONS 6 and 7 to compute values of dipole moment μ and molecular volume V is clearly not legitimate. One might instead assume that the modified equations for $\alpha \neq 0$ describe a heterogeneous system of dipoles,

all having the same moment but with a distribution of effective volumes, a mean value of the latter being determined by the frequency ν_c of maximum absorption. If this is done, one obtains somewhat smaller moments, of the order 3-5 Debye units, and very much larger molecular volumes ranging from 700 to 1000×10^{-24} cm.³

The dubious nature of values obtained for μ and V is apparent from the assumptions. The very broad spread of the $\Delta\epsilon''$ curves evidently could result only from a correspondingly great range in values of the molecular or ionic parameters, and it is not obvious that all species would have closely similar dipole moments. A further difficulty lies in the shape of the loss curves, which, as far as can be judged, have an essentially logarithmic symmetry about the critical frequency. This requires that the contribution of species having individual critical frequencies above and below the observed ν_c be essentially logarithmic also, as a result of dipole moment and size variation. For $\alpha = 0.5$, the necessary range of these frequencies to include 95 per cent of the polarization covers a factor of 10^6 , a difficult number to explain by size variation. The model seems hardly to describe the present situation, and, pending a more adequate theoretical background, a rather qualitative approach here seems to be all that is justified.

It seems reasonable to suppose that very low values of critical frequency in a solvent of low viscosity like benzene must be attributed to strong and extensive interactions among ions, the extreme case of such interaction being the formation of stable compound ions or neutral molecules. Other things being equal, the numbers of such ions which are held together as single, more or less well-defined polar units must be larger to give lower critical frequencies. The existence of broad absorption regions, expressed by large values of the parameter α , for example, must correspond to increasing departures from the idealization of a single homogeneous species. On this basis, it is of interest to examine the data for the various anion substitutions in the octadecyltributylammonium series.

The data for the iodide of FIGURE 6 show a progressive decrease in critical frequencies with increasing concentration, thus indicating a rapid growth in aggregation. The breadth of the curves corresponds to values of α from 0.45-0.55, as found from EQUATION 8, with no obviously great increase with concentrations above 1×10^{-2} .

Unfortunately, the lower range of concentrations ($c = 2 \times 10^{-3}$ to 10^{-2}) for which dipole formation begins has not been investigated for these salts. It would be of interest to trace, for this or a similar system, whether the more complex absorption develops gradually or abruptly from the simple behavior. Presumably, in this range, only the low-frequency limiting slope would be evident from radio-frequency data, but even this should give worthwhile information.

The data for the nitrate, plotted in FIGURE 6, are similar to the iodide results except that the critical frequency for a given salt concentration is

roughly twice as great. Similar comparisons can be made for formate, thiocyanate, and chloride ions, as shown in TABLE 3. From these data, it

TABLE 3
CRITICAL FREQUENCIES AND DISPERSION PARAMETERS FOR OCTADECYLTRI-*n*-BUTYL-AMMONIUM SALTS IN BENZENE

(Frequencies ν_c in megacycles. Values of α calculated from $\Delta\epsilon''(\nu_c)$ using $\epsilon_\infty = 2.272$ and $\epsilon_0 = \epsilon$ of solution at 7.5 kc.

Anion	NO ₃	OOCH	I	NCS	Cl
$\nu_c (c = 1.4 \times 10^{-2})$	50*	50*	25	11	40
ν_c ratio to Cl	1.3	1.3	0.6	0.3	1.0
α	0.50	0.43	0.53	0.57	0.49
$\nu_c (c = 1.7 \times 10^{-2})$	30	—	8.3	—	14.3
ν_c ratio to Cl	2.1	—	0.6	—	1.0
α	0.53	—	0.56	—	0.43

* By Extrapolation.

would appear that less extensive interaction effects occur for the large nitrate ion and the formate ion, while a greater effect occurs for the thiocyanate ion than for the simple chloride and iodide ions. In all cases, aggregate volumes V , however estimated, come out much larger than are reasonable for simple ion pairs, and the values of α obtained fall in the range 0.4–0.6, indicating no single simple species.

Effects of Positive Ion Substitutions. The effect of different cations has already been illustrated for picrates, under the heading, "Calculation of Dipole Movements," where it was seen that there is a very marked difference in behavior of tri-isoamyl and substituted quaternary butylammonium ions, but that addition of long-chain groups in the latter case did not change the qualitative form of the loss curves.

Data for a second series of cation substitutions, carried out for the small symmetrical thiocyanate ion,¹¹ are plotted in FIGURE 7. As would be expected from the other results, the frequencies of significant absorption are very low and the absorption is spread out over a broad range of frequencies. The values of ν_c and the parameter α , calculated as before, are listed in TABLE 4 for a concentration $c = 1.4 \times 10^{-2}$ molar. The substitution of one octadecyl chain lowers ν_c by about a factor of ten and the second by a further factor of seven, and the necessary values of α to fit the data are again of the order 0.4–0.6. The already low values of ν_c and the large changes must certainly indicate extensive interaction effects, while the nature of the curves precludes their being interpreted simply in terms of single species.

Data obtained at other concentrations show appreciably increased equivalent loss factors at higher concentrations, indicating a greater proportion of large aggregates.

The conclusion as to size and complexity are in general agreement with freezing point and conductance data. The former yield polymerization

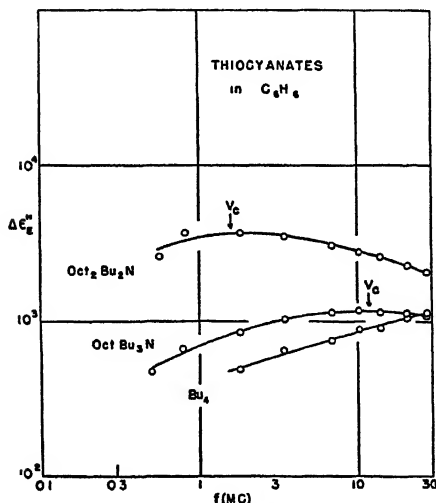


FIGURE 7. Equivalent absorptions vs. frequency for substitution of octadecyl chains in ammonium thiocyanates

numbers increasing rapidly with concentration, which are, moreover, much greater in the octadecyltri-*n*-butylammonium than in the tetra-*n*-butylammonium salt. Conductance measurements¹² show a rapid drop for concentrations greater than the "break" value, indicating a rapid decrease in either number or mobility of ions or both, and this is just as the absorption data indicate.

Related Effects

Relations between absorption measurements and evidence from conductance, freezing point, and low-frequency polarization data have been indicated in the preceding discussion. In addition to these, the observation of non-ohmic field effects on conductance give further support to the general picture, and investigations of dielectric constants at audio-frequencies suggest further complexity of many systems.

Wien Field Effect. Significant increases in low-frequency (60 c.p.s.) conductances with field strength were observed in some of the investigations discussed here. These were found for dilute solutions of low conductance, for which large fields (up to 10 kv/cm.) were necessary to obtain reasonable heat effects as calibration for the absorption measurements in the megacycle range. Because of their incidental nature for the primary objective, only a few measurements were made, which sufficed to establish existence of the effect and give qualitative values. These data indicated little change of conductance for fields less than about 5 kv/cm. At higher fields, an increase became evident which approached a linear change with field strength.

The observations are evidently to be accounted for primarily as a dissocia-

tion of loosely bound aggregates, as treated by Onsager¹² for the case of ion pairs. The theory predicts a linear increase of conductance with field strength for fields great enough that influence of the ion atmospheres can be neglected, the calculated increase being 3.2 per cent per kilovolt/cm. if the conductance is evaluated from calorimetric data. The observations showed comparable but somewhat greater increases. The discrepancies found are not surprising, as the other effects of aggregation all point to much more complex behavior than ion-pair formation in the cases observed. The experimental values did not agree too well between runs with different cells for the unfavorable conditions when the effect was observed. The existence and magnitude of the effect do, however, confirm the other forms of experimental evidence.

Relation to Dielectric Constant Data. The data obtained with the calorimetric method on dipole absorption show behavior of the systems investigated only for a limited frequency range (100 kc. – 28 mc.), and the possibility often exists that a given system may possess other absorption regions either at much higher or lower frequencies, which are missed entirely. The most desirable evidence on this point would obviously be essentially complete frequency coverage from optical frequencies down to sub-audio frequencies, a goal not likely to be soon achieved. Failing this, there exist other checks which may, in some cases, suffice to establish the adequacy of existing results, and in others give some quantitative indication of what is missing.

The basic consideration which makes checks on the data possible is the fact that absorption of energy and dispersion of the dielectric constant over the frequency range are two aspects of the same dipole (or other) polarization. As such, a complete knowledge of the spectrum of one defines the other by well-known integral relations.⁶ As a special case, the area under an absorption curve of $\Delta\epsilon''$ versus frequency is expressed in terms of the limiting dielectric constants ϵ_0 and ϵ_∞ as

$$\int_0^\infty \Delta\epsilon'' d \ln \nu = \frac{\pi}{2} (\epsilon_0 - \epsilon_\infty), \quad (9)$$

a relation readily verified for the specific formulas, EQUATIONS 1 and 5.

If the values of ϵ_0 and ϵ_∞ are known, the form of the absorption curve is then limited by EQUATION 9. In simple cases of absorption, this gives considerable assurance that only one region of significance occurs. In the cases discussed in the "Discussion" section, only a small fraction of the absorption curve was observed, but the fact that dipole moments calculated from its slope and from the difference ($\epsilon_0 - \epsilon_\infty$) agree well with moments computed from the low-frequency dielectric constant and the same polarization model is good indication that no other absorption region of importance

exists in these cases. This agreement, of course, does not preclude the existence of sub-audio frequency effects, but no good reason from equilibrium data is apparent for such effects.

If the absorption behavior is less simple, the situation is not as clear-cut. In the case of the picrate salts for which evidence of two distinct absorption regions was found, the difference $\epsilon_0 - \epsilon_\infty$ of audio-frequency and solvent dielectric constants can only limit the sum of the areas under the two absorption curves. Even this is true only if there is no further absorption at frequencies higher than the one for which ϵ_0 is obtained for the picrates and out of the range of the absorption data.

The situation for the picrates can be better defined experimentally only by further data. The approach which appears most promising is investigation of the dielectric constant ϵ' at frequencies from 10 kc. to 10 mc. or higher. Measurement of ϵ' rather than $\Delta\epsilon''$ appears preferable as, in this range, values of $\Delta\epsilon''$ are obtained as a small difference of large numbers. Together with present absorption data, such measurements should clarify matters greatly.

The data for salts which show broad absorption regions present a special problem, because, as yet, no adequate theoretical model has been devised. The integral relation, EQUATION 9, is unaffected by this, as its validity requires only that the phenomenon be a linear one, but its application fails to give a definite answer if only a small part of the absorption range is measured.

If an empirical equation such as EQUATION 5 is found to describe the measured absorption data when α is chosen to fit any point, the assumption that the rest of the curve is also fitted is reasonable. Tests of this kind have been made on much of the data by calculating α from the observed $\Delta\epsilon''(\nu_c)$ by means of EQUATION 8. The most reasonable available values for ϵ_0 are the highest audio-frequency measurement at 7.5 kc., and ϵ_∞ has been taken to be the dielectric constant of the solvent. The proper values of the latter are, of course, for the solution at sufficiently high frequency ($\nu \gg \nu_c$). These are not available, and the solvent value has been chosen as certainly not greatly in error for small concentrations of solute.

The calculated values of α from EQUATION 8 for various salts and concentrations have been listed in TABLES 3 and 4. Values of α may also be

TABLE 4
CRITICAL FREQUENCIES AND DISPERSION PARAMETERS FOR THIOCYANATES IN BENZENE
($c = 1.4 \times 10^{-2}$, OTHER QUANTITIES AS IN TABLE 3)

Cation	Bu ₄ N*	OctBu ₃	Oct ₂ Bu ₂
ν_c	120	11	1.5
α	0.4	0.57	0.45

* Values estimated from data for $c = 0.936$ and 1.75×10^{-2} .

selected to give the best average fit to observed points. For octadecyltri-*n*-butylammonium nitrate ($c = 2.54 \times 10^{-3}$), the value of α from TABLE 3 is 0.47, and the best fit to the data of FIGURE 6 was found for $\alpha = 0.42$. Similar agreement in the other cases of the "Discussion" Section give support to an assumption that only one, albeit very broad, absorption region exists above audio frequencies.

A somewhat dubious indication of value as to the total absorption can be obtained by assuming that the mechanism of absorption is such that the complex polarization[~] of the Debye theory is governed by the factor $1 + (i\nu/\nu_c)^{1-\alpha}$ rather than $(1 + i\nu/\nu_c)$ and that all aggregates have the same dipole moment. This value can then be computed from either low-frequency data for ϵ_0 or from the absorption data for $(\epsilon_0 - \epsilon_\infty)/\nu_c$ and α . The two results for μ in several cases tried showed agreement to about 20 per cent and further varied in the same ratio with concentration. Such agreement hardly justifies the assumptions, but rather indicates in a different way than before that only one absorption region above audio frequencies of any importance exists.

Low-Frequency Dispersion. The evidence so far discussed does not preclude the existence of further absorption regions at audio and sub-audio frequencies. These, of course, would have to be attributed to extremely sluggish polarizations as a result of very large aggregates or otherwise. Some evidence from dielectric constant data at audio frequencies appears on

* The term here refers to the field vector, not to the additive function of Debye's theory.

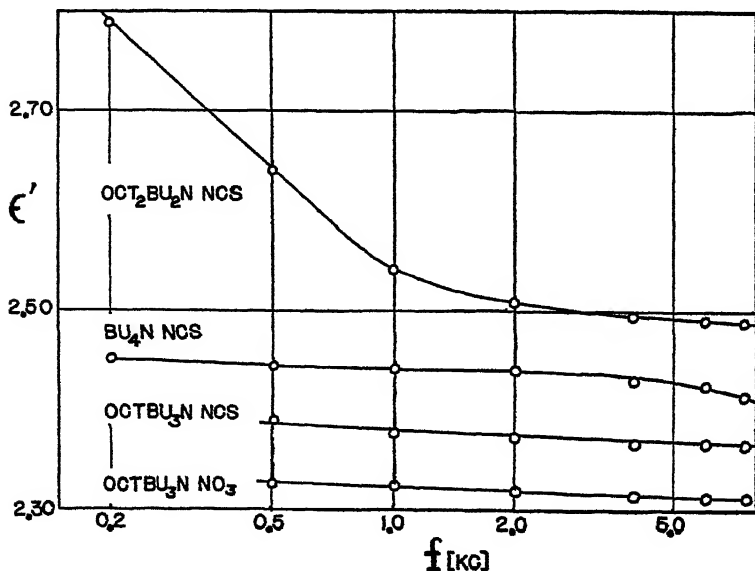


FIGURE 8. Apparent audio frequency dielectric constants of quaternary ammonium salts in benzene.

its face to confirm this. Representative data of ϵ' for several systems, at concentrations of 1.4 to 1.7×10^{-2} , are plotted in FIGURE 8. Only slight variations of ϵ' are indicated for the octadecyltri-*n*-butylammonium nitrate, and the thiocyanate shows only a slight indication of a drop above 3 kc. The dioctadecyldibutylammonium thiocyanate data, however, show a marked rise at low frequencies. A similar behavior has been observed in other solutions at the higher concentrations investigated, the rise at the lowest frequencies being much more marked in some cases.

The reality of this effect has unfortunately not been established with any assurance. The difficulty is, of course, the possible existence of large electrode polarization effects, which can lead to just the observed sort of variation of apparent dielectric constant. The polarization effect is most pronounced for solutions which are good conductors, but little correlation is found between the rise in ϵ' and the audio-frequency conductance. Experimentally, what is wanted is, clearly, measurements using cells with variable spacing of electrodes, in order that the polarization error may be determined, and, if this is not the whole of the effect, corrections to find the true variations in ϵ' can be made.

That the effect may, in fact, be present is at least plausible on comparison with the absorption results at radio frequencies. The observed dispersion effects are most pronounced for systems in which the absorption is large and spread over a wide frequency range, indicating directly extensive and complex interaction. The existence of still more sluggish polarizations is then more plausible in these cases than when the existence of smaller, better defined aggregates is indicated, and the correspondence between the two groups of data from this viewpoint is interesting. Pending experimental data of known significance, much further discussion of such polarizations, except as possibilities not to be excluded without proof, is hardly warranted.

Summary

Measurement of dielectric loss makes possible considerable progress in the understanding of electrolytic systems for which association of ions into polar aggregates occurs. The investigations on which the present discussion is based have shown clearly the difference between systems in which only simple dipole ion-pairs are formed, and cases in which more complex types of aggregation occur. The conclusions drawn from absorption measurements are generally consistent with available conductance, dielectric constant, and cryoscopic data.

The characteristic features of absorption results are that the effects result from polar groups, and that, in contrast with equilibrium measurements, direct evidence of the kinetic behavior of these groups is obtained from the frequency variation. As a result, some idea of the extent and complexity of aggregation can be obtained at a single concentration. Deductions as to the significance on a molecular scale of the results become less quantitative in

more complicated systems because of limited frequency range of measurement, and because of the absence of adequate models.

The future possibilities of dielectric absorption measurements appear to be very great, particularly if their range can be extended to give a more complete spectrum and studies also made of the effect of temperature. With such extensions of experimental methods, a wider range of systems can be examined and a more comprehensive description of the less simple systems obtained.

Bibliography

1. FUOSS, R. M. 1935. Chem. Rev. 17: 27.
2. KRAUS, C. A. 1938. J. Franklin Inst. 225: 687.
3. DEBYE, P. 1929. *Polar Molecules*. Chemical Catalog Co. New York.
4. VAN VLECK, J. H., & V. F. WEISSKOPF. 1945. Rev. Mod. Phys. 17: 227.
5. KIRKWOOD, J. G., & R. M. FUOSS. 1941. J. Chem. Phys. 9: 329.
6. COLE, K. S., & R. H. COLE. 1941. J. Chem. Phys. 9: 341.
7. DEBYE, P. 1934. Phys. Z. 35: 101.
8. HARMS, F. 1901. Ann. Phys. 5: 564.
9. SHARBAUGH, A. H., JR., C. SCHMELZER, H. C. ECKSTROM, & C. A. KRAUS. 1947 J. Chem. Phys. 15: 47.
10. SHARBAUGH, A. H., JR., H. C. ECKSTROM, & C. A. KRAUS. 1947. J. Chem. Phys. 15: 54.
11. STROBEL, H. A., & H. C. ECKSTROM. 1948, J. Chem. Phys. 16: 817.
12. STRONG, L. E. 1941. Thesis, Brown University.
13. ONSAGER, L. 1934. J. Chem. Phys. 2: 599.

FACTORS INFLUENCING THE PROPERTIES OF LONG-CHAIN ELECTROLYTES IN WATER AND WATER MIXTURES OF ORGANIC SOLVENTS

BY PHILIP F. GRIEGER

Lalor Foundation Fellow, Brown University, Providence, Rhode Island

Introduction

This paper is a brief summary of some of the experimental work on long-chain electrolytes that has been done in this laboratory during the past few years.¹ The present experiments were primarily conducted to explore the range of phenomena exhibited by long-chain electrolytes in water-rich solutions. At present, there is no quantitative theory for the facts presented here, and, in the discussion that follows, enumeration of possible explanations of these facts has been avoided, unless, as in the case of the next section, the hypothesis in question admits of direct experimental test.

Concentrations Less Than Critical

Perhaps the simplest hypothesis that has been offered in connection with long-chain electrolytes states that, at concentrations less than critical, these electrolytes are completely dissociated. This hypothesis is consistent with the preponderance of the facts.

In most instances the $\Lambda - \sqrt{C}$ curve can be represented reasonably well by a straight line at concentrations less than critical, and the slope of this line is, in general, of the same order of magnitude as the Onsager slope for a completely dissociated electrolyte.

By and large, such conductance measurements as we have at present have been made with water of specific conductance² 10^{-6} . Since Λ_0 for most of the salts in question is about 100, the solvent correction at $10^{-4} N$ is about 10 per cent, at $10^{-3} N$ about 1 per cent. The theoretical slopes, $d\Lambda/d\sqrt{C}$, are of the order of magnitude 100 so that between $10^{-4} N$ and $10^{-3} N$, say, the decrease in Λ is of the order of 2 Λ units. Then, if, for example, Λ were known precisely at $10^{-3} N$ and with an uncertainty of ± 1 per cent at $10^{-4} N$ (which could easily be the case with a 10 per cent solvent correction), the slope, as determined by these two measurements would be uncertain by $+33$ to -50 per cent. Even if these electrolytes are completely dissociated (at concentrations less than critical), the Onsager slope can hardly be expected to hold within a few per cent at concentrations as high as $10^{-3} N$. Hence, if the Onsager slope is to be taken as a criterion of complete dissociation in these systems, measurements will have to be made with pure water (specific conductance 10^{-7} , or less) down to concentrations $10^{-4} N$, or less. An attempt to make measurements of this sort is now under way by Mr. D. W. Kuhn, of this laboratory.

The present situation in connection with the $\Delta - \sqrt{C}$ slopes is fairly well illustrated by the case of dodecylammonium chloride in water at 25°. Δ for this system* is shown as a function of \sqrt{C} in FIGURE 1. Apparently, Δ is

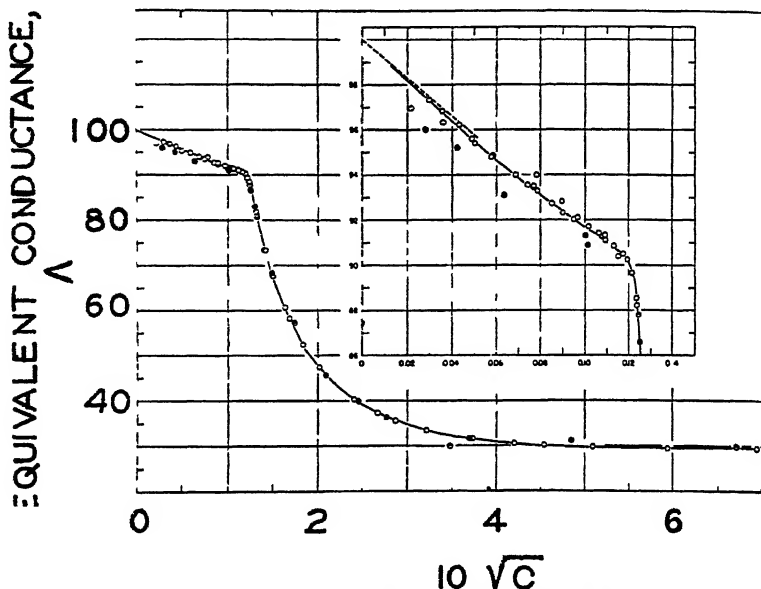


FIGURE 1. Dodecylammonium chloride in water at 25°. Open circles, this laboratory; Solid circles, RALSTON & EGGENBERGER; Half circles, RALSTON, HOERR & HOFFMAN (ref. 3). Dotted line on insert has theoretical slope.

not a linear function of \sqrt{C} up to the break, but shows upward curvature characteristic of strong electrolytes. At concentrations less than $10^{-3} N$, the slope approximates theoretical.

Evidence of a different kind in support of the view that long-chain electrolytes are completely dissociated at concentrations less than critical is shown in FIGURE 2. Here, the distribution of hexadecylpyridonium chloride between water and nitrobenzene is given as a function of water-phase concentration. C , f , and α stand for total salt concentration, mean activity coefficient, and degree of ion-pair dissociation. Primes refer to the water phase; nitrobenzene quantities are unmarked. Curve 1 shows the ratio C'/C ; curve 2, $C'f'/Cf\alpha$. The latter quantity is constant and equal to 0.300 at concentrations less than $5.80 \times 10^{-4} N$ (the critical concentration on the $\Delta - \sqrt{C}$ plot, not shown here) and increases at higher concentrations. $C\alpha f$ is the mean simple ion activity in the nitrobenzene phase, if this solution contains only simple ions in equilibrium with ion pairs. That such is the state of the nitrobenzene phase is indicated by the fact that the Fuoss diagram holds for this solution; the ion-pair dissociation constant is 1.19×10^{-2} .

* The results of RALSTON and co-workers are given for comparison.*

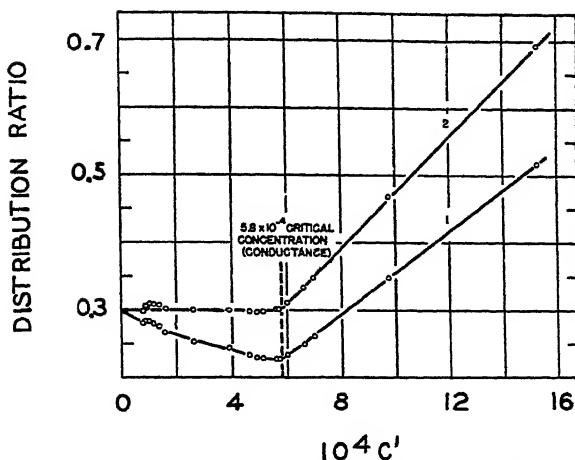


FIGURE 2. Distribution of hexadecylpyridonium chloride between water and nitrobenzene at 25°. Curve 1, c'/c ; Curve 2, $c'f'/c'a$.

Similarly, if the electrolyte is completely dissociated in the aqueous phase, $C'f'$ equals the mean simple ion activity in this solution. Since it is necessary for equilibrium that $a_+ a_- / a_+ a_-$ be constant, the distribution results are consistent with the hypothesis of complete dissociation. These results, of course, apply to water saturated with nitrobenzene (the solubility of nitrobenzene is about 0.2 per cent by weight). The equivalent conductance, for example, of hexadecylpyridonium chloride in pure water and water saturated with nitrobenzene is not the same. In fact, at concentrations greater than critical, the discrepancy is large— Λ is about 20 per cent lower at $10^{-3} N$ in water saturated with nitrobenzene than in water.

Constitution of the Electrolyte

It has been established* that, for a homologous series of long-chain electrolytes, the members of which differ only in the number of carbon atoms in the chain, the critical concentration in water decreases as the length of the chain increases. There are, apparently, no exceptions to this rule. The range in concentration over which the critical concentrations run is large: about 0.4 N for octyltrimethylammonium bromide² down to 8×10^{-5} for octadecyltrimethylammonium oxalate.

No extensive comparison has been made between quaternary long-chain salts with various substituents on the nitrogen atom. It is known, however, that octadecyl and hexadecyl pyridonium salts have lower critical concentrations than the corresponding trimethylammonium salts.

The conductance in water has been determined for at least two diammonium salts having quaternary nitrogens at either end of a ten-carbon

* For references, see McBAIN.⁴

chain. The critical phenomenon is not observed at concentrations up to several times the critical concentration of decylammonium chloride.

Although it is generally recognized that the constitution of the long-chain ion is important in determining the properties of this class of salts, as indicated above, relatively little attention has been paid to the role of the *gegen* ion; it is generally accepted that two long-chain salts which differ only in the identity of the inorganic *gegen* ion will exhibit substantially the same behavior in solution. That this is not always the case, is readily seen from FIGURE 3, where the equivalent conductance is plotted against \sqrt{C} for five

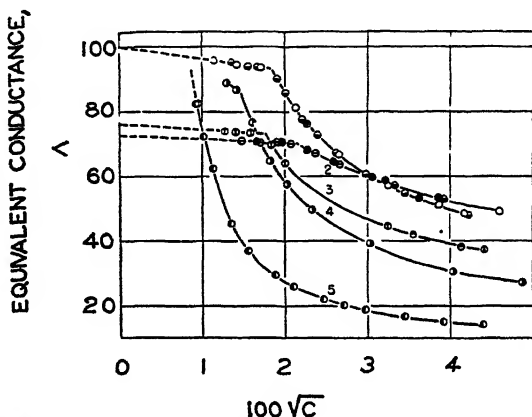


FIGURE 3. Curves for octadecyltrimethylammonium salts in water at 25°. (1) Chloride, (2) formate, (3) bromate, (4) nitrate, (5) oxalate.

n-octadecyltrimethylammonium salts in water. Among the 1-1 salts, the ratio of the extremes in critical concentration is about 2:1, formate compared to nitrate. Even more striking is the fact that the critical concentration of the formate is more than 16 times the critical concentration of the oxalate. Not only does the critical concentration show considerable variation with *gegen* ion for the salts tested, but the slope at the critical concentration approached from the high concentration side shows a marked dependence on *gegen* ion. The magnitude of the slope decreases in general as the critical concentration increases; it is about 10^4 for the oxalate and 10^3 for the formate.

In connection with the maximum phenomenon, the nature of the *gegen* ion makes its most noticeable effect. Thus, octadecylpyridonium nitrate and iodate provide an interesting comparison. The $\Lambda - \sqrt{C}$ plot for the nitrate in water shows the usual breakpoint; in the same solvent, the iodate exhibits a maximum in Λ starting at about $1 \times 10^{-4} N$, the peak is about 10 Λ units high (FIGURE 4, curve 3).

Further comparisons of this sort are possible. Although octadecyltrimethylammonium chloride and nitrate both show the simple breakpoint phenomenon in water, in methanol-water mixtures, containing 10 to 35 per

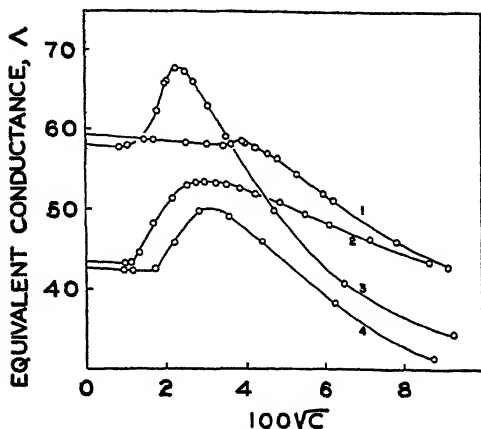


FIGURE 4. Hexadecylpyridonium iodate in: (1) water; (2) 7.94% *t*-butanol. Octadecylpyridonium iodate in: (3) water; (4) 16.18% methanol.

cent methanol by weight, the chloride shows a maximum in Λ , the nitrate does not. The maximum phenomenon is discussed in more detail in the next section in connection with the effect of solvent composition.

Solvent Composition

In pure organic solvents, in general, long-chain electrolytes behave like simple electrolytes. In methanol, for example, the conductance is a linear decreasing function of \sqrt{C} ; the slope conforms closely to theoretical. It is of interest to see how the addition of such substances influences the characteristic behavior of long-chain salts in water.

On the addition of methanol, the conductance of the octadecyltrimethylammonium salts in water is altered in either one of two ways. First, the $\Lambda - \sqrt{C}$ curves show the breakpoint phenomenon at low concentrations of methanol. The breakpoint moves to higher concentration as the methanol content increases; concomitantly, the breakpoint becomes less sharp. In water and mixtures of low alcohol content, the slope $-d\Lambda/d\sqrt{C}$ changes nearly discontinuously at the breakpoint; in mixtures of higher alcohol content (between about 30 and 50 per cent methanol), the slope is continuous at all salt concentrations; above about 50 per cent methanol, Λ is essentially rectilinear in \sqrt{C} . In this class belong the nitrate, oxalate, and bromide. The $\Lambda - \sqrt{C}$ curves for the oxalate, which illustrate these points, are shown in FIGURE 5.

On the other hand, the conductance plots of the chloride, bromate, iodate, and formate all exhibit maxima in mixtures that contain between about 10 and 35 per cent methanol. The peak in Λ is highest at about 20 per cent alcohol. The height of the peak in 20 per cent methanol decreases as the *gegen* ion is iodate (4 units), formate (3.5 units), bromate and chloride (2

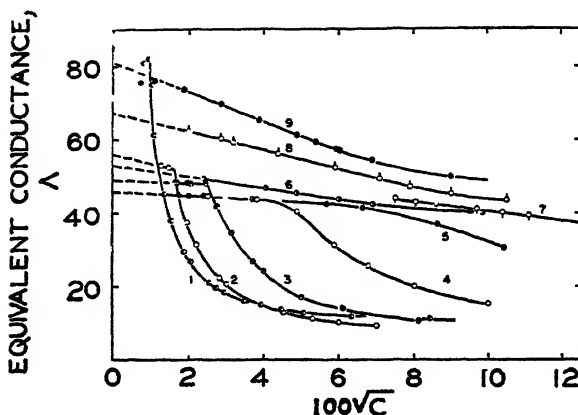


FIGURE 5. Octadecyltrimethylammonium oxalate in water-methanol mixtures. Wt. % methanol: (1) zero, (2) 20.7, (3) 30.2, (4) 40.1, (5) 50.3, (6) 69.8, (7) 75.0, (8) 89.8, (9) 100.00.

units). Curves for the chloride in 0 to 45 per cent methanol are shown in FIGURE 6. As may be seen from this figure, the maximum phenomenon becomes less pronounced as the methanol content is increased beyond 20 per cent and finally disappears. For mixtures in the range 50 to 100 per cent, the curves resemble those of the oxalate.

Results of a similar nature obtain in acetone-water mixtures. The oxalate shows the breakpoint phenomenon. The bromate and formate show the maximum effect just as they do in methanol-water mixtures, but

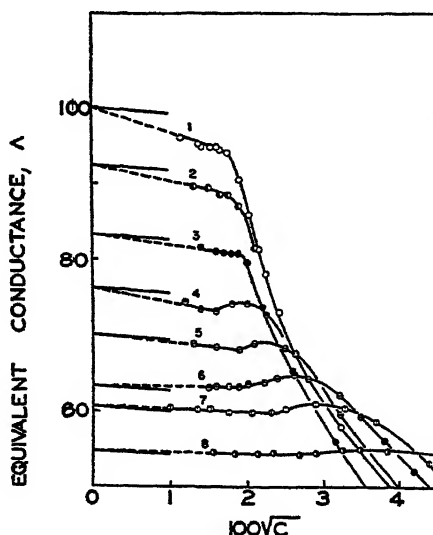


FIGURE 6. Octadecyltrimethylammonium chloride in water-methanol mixtures at 25°. Wt. % methanol: (1) zero, (2) 2.3, (3) 4.7, (4) 10.1, (5) 12.9, (6) 17.7, (7) 22.1, (8) 30.1. Solid lines at left have Onsager slope.

the height of the peaks is considerably greater (by a factor of two or three) in acetone-water mixtures, than in methanol-water mixtures.

A fairly good correlation between the critical concentration in methanol-water mixtures on the one hand, and acetone-water mixtures on the other, and the solvent dielectric constant exists for octadecyltrimethylammonium oxalate. The two curves for critical concentration against $1/D$ nearly coincide; the relationship is approximately exponential. The good correlation between critical concentration and D in the case just mentioned may be fortuitous. In any event, it is certainly not general. In the case of dodecylammonium chloride, as will be pointed out later, such a simple relationship clearly does not appertain. Measurements with higher type alcohol-water mixtures would be of interest in connection with the oxalate.

At the present time, we have only a few measurements on octadecyl and hexadecylpyridonium iodate in alcohol-water mixtures. Both these salts show conductance maxima in pure water. The addition of 8 per cent tertiary butanol in the case of the hexadecylpyridonium salt increases the height of the maximum about tenfold and markedly changes the form of the conductance curve. In 16 per cent methanol, the octadecylpyridonium salt shows roughly the same conductance curve that it does in water. These curves are shown in FIGURE 4.

With dodecylammonium salts in methanol-water, *isopropanol*-water, and *t*-butanol-water mixtures, only the break-point phenomenon is found. Unlike the case of the octadecyltrimethylammonium compounds mentioned before, the addition of organic solvent does not always produce an increase in critical concentration. The present results may be summed up by saying that the higher (more C atoms) the alcohol, the more pronounced is the tendency to decrease the critical concentration (FIGURE 7). It is clear that,

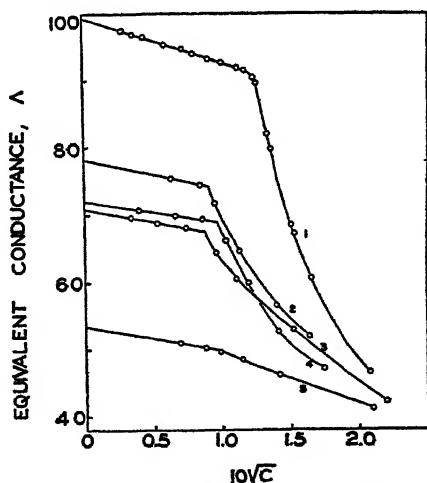


FIGURE 7. Dodecylammonium chloride in water-alcohol mixtures at 25°. (1) Water, (2) 5.7% *t*-butanol, (3) 7.9% *t*-butanol, (4) 8.0% *i*-propanol, (5) 16.2% *i*-propanol.

for dodecylammonium chloride in alcohol-water mixtures, the critical concentration is not determined solely by solvent dielectric constant. It may be noted that the depression of the critical concentration is greatest at low alcohol content.

In connection with the maximum effect, it may be noted also that, all other things being equal, the peak is greater the longer the chain (octadecylpyridonium iodate, 10 units *vs.* hexadecylpyridonium iodate, 1 unit in water); *i.e.*, the tendency to exhibit a maximum increases as the critical concentration decreases as far as variation in chain length goes. On the other hand, the tendency to form a maximum varies with *gegen* ion in the opposite sense, that is, the maximum in methanol-water mixtures is generally greater the higher the critical concentration in water.

A comparison of the octadecyltrimethylammonium salts in methanol-water mixtures shows that the tendency to form a maximum increases as the critical concentration in water increases, and, roughly, as the mobility of the *gegen* ion decreases, or (roughly) as the size of the *gegen* ion increases.

In addition, for octadecyltrimethylammonium chloride, bromate and formate in methanol-water mixtures, and the bromate and formate in acetone-water mixtures, the rise point comes, in many cases, at a concentration less than the critical concentration in water.

For the sake of completeness, it may be put down that the maximum always comes at the dilute end of the concentration axis. The rise point terminates the linear portion of the conductance curve.

Addition of Other Compounds in Small Amount

The addition of small amounts of dodecanol significantly alters the conductance of dodecylammonium chloride in 25 per cent methanol-water mixture. This is shown in FIGURE 8. The methanol was added to the system

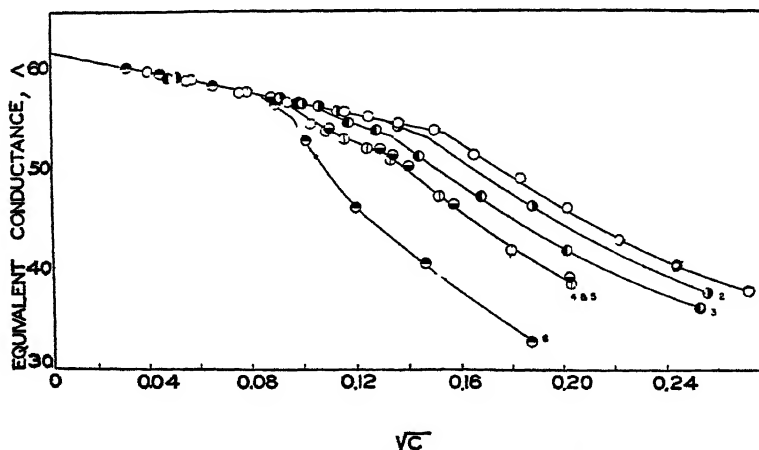


FIGURE 8. Dodecylammonium chloride in 24.65% methanol. Moles dodecanol per mole salt: (1) zero, (2) 0.044, (3) 0.11, (4) 0.25, (5) 0.26, (6) 0.65.

to increase the solubility of dodecanol; in spite of this, the system was cloudy over the entire range of concentration at mole ratio of dodecanol to salt 0.65 and was cloudy over part of the concentration range at mole ratio 0.25.

Two features of the curves, at least, are noteworthy: (1) at concentrations less than critical the curves coincide; (2) the critical concentration decreases with increase in dodecanol content. The first point is in agreement with the proposition that long-chain electrolytes are completely dissociated at concentrations less than critical. It will be noted that the change in critical concentration on addition of dodecanol is of an entirely different order of magnitude than that which obtains on addition of alcohols of lower type (those containing one to four C atoms per molecule). In a sense, this result is consistent with the rule given above, that, the greater the number of C atoms in the alcohol, the greater the tendency to depress the critical concentration. In one important respect, the case under discussion is much simpler than the cases given in the preceding section; the change in critical concentration is not accompanied by a large decrease in solvent dielectric constant.

The curves appear to possess two break points. The curves, as drawn, overemphasize this feature perhaps, but there seems to be a pretty clear change in slope in several of the cases that approximates the magnitude of the change in slope that occurs on the straight-line portion of the curve. This feature deserves further study.

The addition of common ion inorganic electrolytes, in general, depresses the critical concentration and suppresses the conductance maximum. In all cases tested, the conductance of a mixture, such as potassium chloride and octadecyltrimethylammonium chloride, is less than the sum of the conductances of the separate salts.

References

1. BROWN University Theses by E. C. EVERS (1941); P. F. GRIEGER (1947); G. L. BROWN (1947); H. S. YOUNG (1948). E. C. EVERS, P. F. GRIEGER & C. A. KRAUS. J., *Am. Chem. Soc.* **68**: 1137; G. L. BROWN, P. F. GRIEGER, E. C. EVERS, & C. A. KRAUS. 1947. *Ibid.* **69**: 1835; E. C. EVERS & C. A. KRAUS. 1948. *Ibid.* **70**: 3049; P. F. GRIEGER & C. A. KRAUS. 1948. *Ibid.* **70**: 3803; G. L. BROWN, P. F. GRIEGER, & C. A. KRAUS. 1949. *Ibid.* **71**: 95; H. S. YOUNG, P. F. GRIEGER, & C. A. KRAUS. 1949. *Ibid.* **71**: 309.
2. SCOTT, A. B., H. V. TARTAR, & H. C. LINGAFELTER. 1943. *J. Am. Chem. Soc.* **65**: 692. (Used water of specific conductance less than 10^{-6} , but they do not state what the conductance was.)
3. RALSTON, A. W., C. W. HOERR, & E. J. HOFFMAN. 1942. *Ibid.* **64**: 97. RALSTON, A. W., and D. N. EGGENBERGER. 1948. *Ibid.* **70**: 436.
4. ALEXANDER's *Colloid Chemistry*. 1944. V: 102. Reinhold Publishing Company, New York.

THE VISCOSITY OF MIXTURES OF POLYELECTROLYTES AND SIMPLE ELECTROLYTES*

BY RAYMOND M. FUOSS AND ULRICH P. STRAUSS†

Sterling Chemistry Laboratory, Yale University, New Haven, Connecticut

Introduction

Viscosity‡ has been a subject of active interest to polymer chemists ever since Staudinger² proposed that the intrinsic viscosity was a measure of molecular size. Many years before, Einstein³ had shown that the specific viscosity of a suspension of spherical particles was proportional to their volume fraction, regardless of size. In contrast, Staudinger found that the same concentration of different polymers gave different specific viscosities, which increased with increasing molecular weight. Staudinger's original hypothesis of direct proportionality between limiting reduced viscosity and molecular weight has subsequently been modified, but a one-to-one correspondence, in accordance with a linear log-log relationship, between the intrinsic viscosity and molecular weight for a homologous series of polymer fractions has been established by numerous investigations covering a wide variety of materials.⁴ Debye⁵ has recently given a simple theoretical derivation of Staudinger's rule and has shown that the latter corresponds to an idealized model of the polymer chain. Real polymers can be shown⁶ to deviate from the model in a way which can be correlated with the empirical function just mentioned.

For a given polymer, the reduced viscosity is, at low concentrations, usually a linear function of the concentration

$$\eta_{sp}/C = [\eta] + k'[\eta]^2C \quad (1)$$

where k' equals approximately 0.38, apparently independent of the polymer. Debye⁷ interprets the intrinsic viscosity as a measure of the volume which the polymer molecule occupies as a loosely coiled chain in solution. Some early observations of Staudinger and Trommsdorff⁸ and by Kern⁹ on the sodium salt of polyacrylic acid and some recent work of ours¹⁰ on poly-4-vinyl-*N-n*-butylpyridonium bromide (P-ViBuPyBr), a polyelectrolyte of quite different structure, resulted in viscosity curves which were completely unlike that given by the function of EQUATION 1. Instead of decreasing linearly to a limiting value with decreasing concentration, the reduced viscosity of the above polyelectrolytes increases, apparently without limit

* This research was partially supported by Contract N6 ori 44-T.O.4 with the Office of Naval Research
† Sterling Research Fellow, Yale University, 1946-48.

‡ We shall use, with one exception, the nomenclature, units, and symbols proposed by L. H. CRAIG,¹ according to which the specific viscosity η is the ratio $(\eta - \eta_0)/\eta_0$ for a solution of concentration C grams of solute per 100 cc. of solution, which has a viscosity η in a solvent of viscosity η_0 . The relative viscosity is the ratio η/η_0 . The reduced viscosity is defined as η_{sp}/C and the intrinsic viscosity $[\eta]$ is the limit of η_{sp}/C as C approaches zero. We shall reserve the symbol c for concentration expressed as equivalents per liter.

on a linear scale, with decreasing concentration, along a curve which is strongly concave upwards. Qualitatively, one might assume that dissociation of the polyelectrolyte increases with dilution, and, due to increasing Coulomb repulsion between the increasing number of unpaired residual charges on the chain, the polyion coil expands and thus acquires a steadily larger reduced viscosity. This effect is undoubtedly present, but we do not believe that the whole story is quite so simple, because interionic forces both between polyions and polyions and between polyions and the accompanying counter ions must also be taken into account.

The purpose of the research reported here was to obtain quantitative data of reasonable precision on a representative polyionic system. The viscosities in water of an unfractionated sample of poly-4-vinyl-*N-n*-butylpyridonium bromide of average molecular weight 500,000 have been measured over the concentration range from 0.003 to 1.0 g./100 cc. solution, at several rates of shear. Empirically, we can represent the data by an equation of the form

$$\eta_{sp}/C = a/(1 + b\sqrt{C}) \quad (2)$$

(although we suspect the presence of an additive constant on the right, as will be brought out in the discussion).

In order to enhance the interionic effects, we have also measured the viscosities of the polysalt in the presence of added simple electrolytes such as potassium bromide. Addition of excess simple electrolyte decreases the viscosity of the polyelectrolyte markedly, first producing a maximum in the viscosity-concentration curve, and eventually permitting extrapolation according to EQUATION 1. The limiting values thus obtained depend on the ionic strength μ of the simple electrolyte present in the solvent medium according to the empirical equation

$$[\eta] = A(1 + B/\sqrt{\mu})^2. \quad (3)$$

The constant a of EQUATION 2 is the limiting value of the reduced viscosity at zero ionic strength* and is several orders of magnitude larger than the constant A of EQUATION 3 which is the limit corresponding to infinite ionic strength. EQUATIONS 2 and 3 are, of course, limiting forms of the same (as yet unknown) function. We are unable to suggest its form at present because we do not know how to compute the contribution of the polyion to ionic strength.

Apparatus and Materials

Viscometers. Viscometers were constructed according to Bingham's design¹¹ and were operated at constant pressures, ranging from 30 to 200

* More accurately, "at very low ionic strength," i.e., that of pure water, where $[H^+] = [OH^-] = 10^{-7}$.

cm. of water. Corrections* for kinetic energy and drainage were made, based on calibration runs made with glycerol and sucrose solutions which presumably do not exhibit structural viscosity. The corrected pt (p = pressure, t = time of flow) values still depended on driving pressure. We ascribe this change to a structural viscosity which depends on rate of shear and shall discuss this point in a later section of this paper.

Viscometer *A* was constructed of lead glass, with a capillary bore of 0.0148 cm.; viscometer *B* (working volume, 4.045 cc.) was of pyrex with a capillary bore of 0.0189 cm. and length of 11.0 cm. For moderately concentrated solutions, the two viscometers gave concordant results, as shown in TABLE 1, where reduced viscosities are given for two aqueous solutions of

TABLE 1
COMPARISON OF VISCOMETERS

C	Viscometer	$\beta = 2000 \text{ sec}^{-1}$			$\beta = 3500 \text{ sec}^{-1}$		
		p	t	η_{sp}/C	p	t	η_{sp}/C
0.2321	<i>A</i>	87.90	530.0	4.98	150.20	304.0	4.81
0.2353	<i>B</i>	71.00	255.0	4.96	121.90	146.0	4.80

PViBuPyBr of nearly the same concentration at two velocity gradients. In a capillary viscometer, the velocity gradient varies with distance from the center; an average may, however, be calculated¹³ by Kroepelin's formula:

$$\beta = 8V/3\pi r^2 t \quad (4)$$

where V is the volume flowing through a capillary of radius r in t seconds. In TABLE 1, pressures are given as g./cm.², calculated from a measured water head. While it is obvious that the viscosity depends on β , the velocity gradient, it will be noted that, at the same velocity gradient, the two viscometers agree despite the large differences in radius, driving pressure, and time of flow. In more dilute solutions, however, viscosities obtained with Viscometer *A* were found to change with time; at 0.06 g./100 cc., for example, the viscosity decreased by about 5% in 18 hours. Since the pH of the solution in the viscometer rose during this time from 4.3 to 5.6, we assumed that alkali was leaching out of the glass. Our final results on dilute solutions were therefore all determined using Viscometer *B*, in which no change of viscosity with time was observed.

Polyvinylpyridine. 360 cc. of C. P. toluene were boiled to expel air and cooled with a stream of nitrogen bubbling through. Then 40 cc. of freshly

* The usual viscometer formula is $\eta/K = pt - \lambda/t$ where K is the viscometer constant and λ is the coefficient for the kinetic energy correction. In another investigation (G. I. CATHERS¹³), it was found that the drainage error could also be included by writing $\lambda/\rho = A - B(pt/\rho)^2$ where ρ is density, and A and B are constants. Calibration with sucrose and glycerol solutions gave $A = 1.57 \times 10^{-4}$ and $B = 2.92 \times 10^{-8}$.

giving $pt = 9371$. The correction coefficient λ then equals 1.31×10^4 and the correction to the pt product for kinetic energy and drainage equals 138. From $\eta/K = 9233$, and $\eta_0/K = 8347$, we obtain $\eta_{sp} = 0.1061$, whence $\eta_{sp}/C = 0.734$.

distilled 4-vinylpyridine (B. P. 72° at 20 mm.) and 1.2 g. of benzoyl peroxide were added, and the mixture shaken in a water bath at 40° for 46 hours, during which time insoluble polymer formed. The solid was centrifuged out, washed twice with benzene, dried in vacuum, re-suspended in benzene and stirred for one hour, and, after centrifuging, was dried in a vacuum oven at 40° for 5 days. The yield was 7.9 g. or 20 per cent.

The viscosity of the polymer was determined in 91.4 weight % alcohol at 25°. The results are given in TABLE 2. The data satisfy EQUATION 1,

TABLE 2
POLYVINYLPIRIDINE IN ETHANOL

<i>C</i>	η_{sp}	η_{sp}/C
0.1306	0.206	1.58
0.253	0.422	1.67
0.388	0.691	1.78

with $[\eta] = 1.48$ and $k' = 0.35$. The specific volume⁷ corresponding to this intrinsic viscosity is 59.2 cc./gm. Osmotic pressure measurements in ethanol gave a (number average) molecular weight of 207,000 (degree of polymerization, approximately 2,000); we are indebted to Dr. W. Albrink for these determinations.

Poly-4-Vinyl-N-n-Butylpyridonium Bromide. 2.88 g. of the above parent polymer were dissolved in 57.3 g. of nitromethane, and 24.2 g. of *n*-butyl bromide were added. The mixture was heated at 55° for four days, after which the solvent and excess butyl bromide were evaporated at reduced pressure. The product was dissolved in ethanol and precipitated in dioxane. After drying under vacuum, bromine was determined by the Parr semi-micro-bomb method: found, 32.2, 31.9, 32.3%; average, 32.1%; calculated, 33.0%.

Experimental Results

The viscosity data are given below in Tables 3-11, where *C* is concentration of poly-salt in grams per hundred cc. of solution and added salt concentrations are given in moles per liter. All data refer to 24.70°. In determining the viscosities of the mixed electrolytes, a salt solution of definite molality was prepared by weight from C. P. grade chemicals, and this solution was then used as solvent for the poly-salt which was added in various amounts. Volume concentrations (molarity) were computed from the known densities. In computing the specific viscosities of the systems in which salt solutions were used as solvents for the poly-salt, the definition of Footnote 1 was used, where η_0 was, of course, the viscosity of the salt solution.

At the higher added salt concentrations, the change of apparent viscosity with rate of shear was less than 0.5 per cent for a change of 1000 sec.⁻¹ in β ,

TABLE 3
VISCOSITIES OF PViBuPyBr IN WATER

C ($\beta =$	η_{sp}			η_{sp}/C		
	2000	3500	5000	2000	3500	5000)
0.00289	0.063 ₈	0.054 ₂	0.048 ₄	22.0	18.8	16.7
.00309	.067 ₁	.060 ₆	.054 ₉	21.7	19.4	17.8
.00872	.152 ₉	.136 ₉	.126 ₄	17.5	15.7	14.5
.00969	.165 ₉	.150 ₅	.139 ₁	17.1	15.5	14.4
.0150	.223 ₈	.202 ₃	.187 ₉	14.9 ₂	13.4 ₈	12.5 ₂
.0263	.330	.303	.282	12.5 ₃	11.5 ₆	10.7 ₃
.0564	.516	.483	.456	9.15	8.56	8.08
.0592	.531	.500	.470	8.98	8.45	7.94
.1103	.762	.723	.687	6.90	6.55	6.23
.2321	1.159	1.116	—	4.98	4.81	—
.2353	1.168	1.130	1.096	4.96	4.80	4.66
.6936	2.129	2.103	—	3.07	3.03	—
.9875	2.593	2.567	—	2.63	2.60	—

TABLE 4
VISCOSITIES OF PViBuPyBr IN DILUTE POTASSIUM BROMIDE SOLUTIONS

$(\beta =$	η_{sp}/C		
	2000	3500	5000)
KBr = 0.001082 M			
0.0281	3.25	3.12	2.98
.0533	3.59	3.46	3.33
.0787	3.71	3.60	3.45
.1225	3.71	3.63	3.50
.2264	3.51	3.44	3.37
.3231	3.25	3.20	—
.7873	2.56	2.53	—
KBr = 0.01029 M			
0.1141	1.25	1.25	1.25
.1146	1.27	1.27	1.27
.2369	1.35	1.34	1.33
.3819	1.38	1.37	1.36
.5059	1.41	1.40	1.40
1.002	1.44	1.43	—

TABLE 5
VISCOSITIES OF PViBuPyBr IN POTASSIUM BROMIDE SOLUTIONS

C	η_{sp}/C	C	η_{sp}/C
KBr = 0.0335 M		KBr = 0.2438 M	
0.1445	0.74 ₈	0.2845	0.33 ₂
.4858	.804	.5784	.340
.8497	.841	.8994	.354
KBr = 0.1003 M		KBr = 1.086 M	
0.1488	0.471	0.348	0.244
.4810	.485	.710	.254
1.090	.520	1.028	.259

TABLE 6
VISCOSITIES OF PViBuPyBr IN DILUTE POTASSIUM SULFATE SOLUTIONS

C ($\beta =$	η_{sp}/C		
	2000	3500	5000)
$K_2SO_4 = 0.0003042 M$			
0.0169	2.24	2.14	2.14
.0313	3.26	3.11	3.00
.0473	3.83	3.66	3.51
.0687	3.24	3.14	3.04
.0787	4.14	3.97	3.84
.0951	4.15	3.98	3.86
.1363	4.07	3.93	3.82
.1576	3.80	3.68	3.59
.3160	3.56	3.49	3.40
.5940	2.90	2.86	—
$K_2SO_4 = 0.0003617 M$			
0.0261	2.52	2.43	2.32
.0472	3.33	3.25	3.16
.0722	3.87	3.73	3.62
.1034	3.83	3.71	3.58
.1109	3.83	3.70	3.59
.1188	3.86	3.75	3.63
.1704	3.83	3.72	3.62
.317	3.43	3.35	3.28
.555	2.91	2.87	—
.887	2.50	2.47	—

TABLE 7
VISCOSITIES OF PViBuPyBr IN POTASSIUM SULFATE SOLUTIONS

C	η_{sp}/C	C	η_{sp}/C
$K_2SO_4 = 0.00303 M$		$K_2SO_4 = 0.09093 M$	
0.1314	1.057	0.1481	0.500
.2788	1.282	.4102	.524
.4586	1.449	.8798	.555
.6073	1.509		
1.171	1.567		
$K_2SO_4 = 0.02895 M$		$K_2SO_4 = 0.3532 M$	
0.2026	0.576	0.3572	0.485
.4147	.598	.6847	.497
.7064	.630	1.067	.512

TABLE 8
VISCOSITIES OF PViBuPyBr IN 0.0002707 M MAGNESIUM SULFATE SOLUTION

	η_{sp}/C		
($\beta =$	2,000	3,500	5,000)
0.0114	1.82	—	1.71
.0270	3.23	3.08	2.95
.0443	3.81	3.65	3.55
.0671	4.21	4.05	3.90
.0968	4.30	4.15	4.00
.1290	4.25	4.13	4.01
.1721	4.09	3.96	3.84
.2637	3.75	3.65	3.57
.4695	3.19	3.13	3.08
.8996	2.53	2.50	

TABLE 9
 VISCOSITIES OF PViBuPyBr IN MAGNESIUM SULFATE SOLUTIONS

<i>C</i>	η_{sp}/C	<i>C</i>	η_{sp}/C
$MgSO_4 = 0.001779 M$		$MgSO_4 = 0.02687 M$	
0.0442	1.13	0.2159	0.619
.0990	1.37 ₂	.4267	.647
.1916	1.62 ₀	.7649	.693
.3054	1.80 ₀	$MgSO_4 = 0.08129 M$	
.4703	1.879	0.1998	.542
.8005	1.857	.4917	.565
$MgSO_4 = 0.008438 M$.8070	.587
0.1707	0.787	$MgSO_4 = 0.2035 M$	
.3627	.873	0.3443	0.521
.6525	.961	.7464	.546
		1.080	.560

 TABLE 10
 VISCOSITIES OF PViBuPyBr IN MAGNESIUM BROMIDE SOLUTIONS

<i>C</i>	η_{sp}/C	<i>C</i>	η_{sp}/C
$MgBr_2 = 0.004072 M$		$MgBr_2 = 0.05251 M$	
0.0418	1.33	0.1576	0.44 ₀
.0801	1.40	.2590	.462
.1445	1.44	.4126	.469
$MgBr_2 = 0.01729 M$		$MgBr_2 = 0.1847 M$	
0.0951	0.73 ₁	0.2170	0.29 ₄
.2343	.74 ₄	.4537	.303
.4014	.77 ₀	.6882	.311

 TABLE 11
 VISCOSITIES OF PViBuPyBr IN 91.4 WT. % ETHANOL

<i>C</i>	η_{sp}/C
0.0636	2.31 ₁
.1158	1.80 ₀
.301	1.27 ₈
.474	1.09 ₀
.782	0.925

and here data for only one driving pressure are given (although others were measured). The notations 2,000, 3,500 and 5,000 at the heads of the various columns refer to $\beta = 2,000 \text{ sec.}^{-1}$, etc.; in some cases, these values were interpolated from plots of η_{sp} versus β , and in others were determined by pre-setting the pressure to give the above round values of β . The latter method was very convenient: A run at an arbitrary pressure was made, and

then, from the resulting pt product, the pressure required to give the flow time which corresponded to the desired value of β (cf. EQUATION 3) was calculated. For Viscometer *A*, $\beta = 1.06 \times 10^6/t$ and for Viscometer *B*, $\beta = 0.509 \times 10^6/t$; for $\beta = 2,000, 3,500$ and $5,000$, the flow times were 530, 303, and 212 sec. for *A* and 255, 146 and 102 sec. for *B*. Corrections for small deviations were naturally made.

Discussion

The electrolytes which we are considering have similarities and differences when considered with respect to ordinary electrolytes, ordinary polymers, and colloidal electrolytes such as proteins and soap micelles. We may, therefore, expect some parallels in properties as well as some differences. Like the colloidal electrolytes of natural origin, these synthetic electrolytes have high molecular weights and high charges; unlike them, however, the linear polyelectrolytes have a readily deformable, very mobile, structure. The configuration of the polyelectrolyte in solution will be sensitive to concentration because the net charge on a polyion will change as ionic association changes with concentration. Then, since the segments of the chain are free to move with respect to each other, the polymer coil should expand as the solution is diluted and increasing intramolecular Coulomb repulsion makes itself felt.

A linear polyelectrolyte can thus accommodate itself to a changing electrical environment by changing its shape, and the potential distribution will change accordingly. It seems likely, therefore, that the concept of electrokinetic potential, which has proven to be so useful in the case of particles whose size and shape remain fixed, will be less valuable in treating the properties of these electrolytes than the admittedly more difficult detailed treatment of charge distribution within as well as around the polycation. At large distances from a coiled polyion, it may well act as a uniformly charged particle, but local interaction cannot be described in terms of a double layer, because the coil presents no fixed surface around which a layer can form. The polyion may perhaps best be compared in its behavior to a charged droplet of concentrated solution with a rather diffuse boundary. In ordinary electrolytes, the charge distribution is nearly uniform in the sense that a microscopic test element of volume will contain, on an average, the same number of positive and negative ions and this number will equal the gross macroscopic concentration. In the chain polyelectrolyte, however, positive ions are constrained to remain near each other by the valence bonds between the atoms, and a microscopic test element will encounter wide fluctuations in charge density; between the positive polyions, it will contain only negative counter ions, while in the vicinity of the polyion it will contain both positive and negative ions, both in amounts larger than the macroscopic average density. As Flory¹⁴ first pointed out in his treatment of the thermodynamic properties of chain polymers, a uniform distribution of

solute, from a molecular point of view, is inherently impossible, especially at low concentrations.

Some of these configuration changes may be deduced qualitatively from curve 1 of FIGURE 1 and the data shown in FIGURE 2. The first curve gives

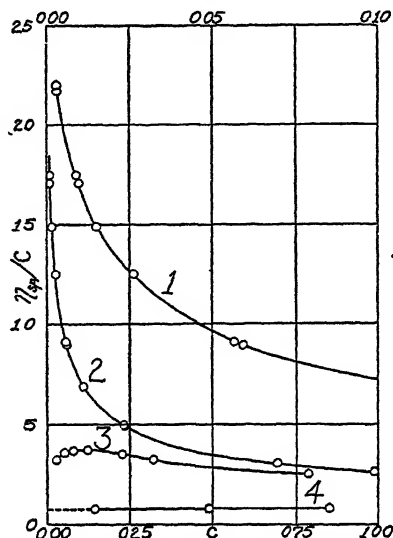


FIGURE 1. Viscosity at $\beta = 2,000$ of polysalt. curve 1, in water, abscissae above; curve 2, in water, abscissae below; curve 3, in 0.001082 *N* potassium bromide solution, curve 4, in 0.0335 *N* potassium bromide solution.

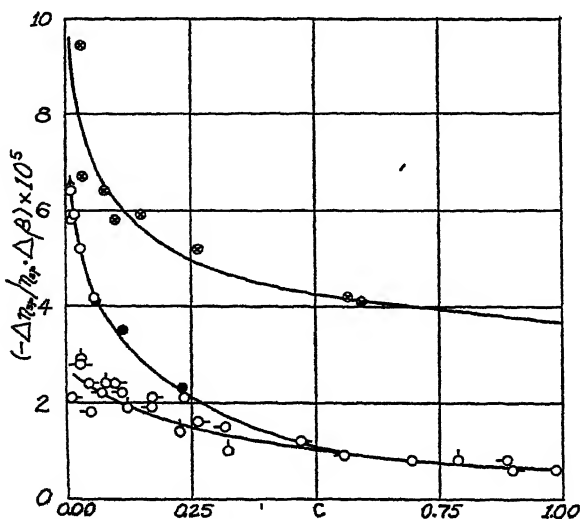


FIGURE 2. Dependence of viscosity on rate of shear. Upper curve, dilute range in water, abscissae magnified ten times. Middle curve, whole range in water; solid circles, viscometer A, open circles, viscometer B. Lower curve, whole range in salt solutions; dashes up, 0.001082 *N* potassium bromide; dashes left, 0.0003617 *M* potassium sulfate; dashes right, 0.0002707 *M* magnesium sulfate.

the reduced viscosity of the polybromide as a function of concentration in the very dilute region (up to 0.1 g. poly-salt per 100 cc. of solution). Ordinary uncharged polymers would show very little change of viscosity with concentration here; we recall the linear form of EQUATION 1 and the magnitude of the two terms: for an intrinsic viscosity of 8, η_{sp}/C would decrease by only about 0.5% on going from 0.1 g./100 cc. to zero, while here the reduced viscosity *increases* by a factor of three in the range from 0.10 to 0.003, the lowest concentration measured. There is no tendency visible of an approach to a finite limiting value on this scale. Furthermore, the reduced viscosity is an order of magnitude larger than that of the parent polymer, for which $[\eta]$ was found to be 1.48. If we assume that the reduced viscosity is a measure of the average volume occupied by the polymer molecule, then we must immediately conclude that the polyelectrolyte has a much more diffuse structure than that of the parent polymer, and, further, that it becomes more open as concentration is reduced. Coulomb repulsion between unpaired* positive charges can account for the increase in the size of the coil over that corresponding to the neutral polymer from which it was made. With increasing dilution, negative ions have increasingly greater probabilities of escape from the polyion, and the resulting increased repulsion between newly uncompensated pyridonium groups would lead to a further increase in size. The limit beyond which the particle cannot increase would, of course, be the completely extended chain, which would correspond to the hypothetical limit of infinite dilution where the polyion would have no associated counter ions at short distances from its positive charges.

As shown in FIGURE 2, the observed viscosities depend on *rate of shear*. Our measurements covered the approximate range $2,000 \leq \beta \leq 5,000 \text{ sec.}^{-1}$. The ordinate is the average chord slope of specific viscosity-velocity gradient plots; these showed a slight curvature in the above range of variables, but for qualitative discussion the chord slope is useful. The units are such that the ordinate gives the per cent change in specific viscosity for a change of $1,000 \text{ sec.}^{-1}$ in β . The top curve is for the very dilute range and the middle curve for the whole range of concentration covered. It will be noted that the dependence of viscosity on shear for the polyelectrolyte in water is rather large and increases with dilution. Since non-Newtonian flow can be produced by asymmetrical particles, we may interpret the curves as meaning that asphericity increases with dilution, in agreement with our argument that the chain uncoils as local charge compensation in the chain is diminished on dilution.

Added simple electrolyte considerably reduces the dependence of viscosity on rate of shear. The lower curve of FIGURE 2 shows the effects of three different salts at the same ionic strength (0.001082 molar potassium chloride, 0.0003617 molar potassium sulfate, and 0.0002707 molar magnesium sulfate).

* We are using "ion-pairs" in the sense of the generalized definition whereby to every positive ion in a solution a negative ion at some distance r may be assigned. Depending on the magnitude of the variables, this distance may be of the order of molecular dimensions or much larger. (Cf. R. Fuoss, *Trans. Faraday Soc.* 30: 967, 1934.)

In these experiments, the named electrolytic solutions were used instead of water as the solvent for the polyelectrolyte; *i.e.*, the concentration of simple electrolyte remained fixed, while that of the polyelectrolyte was varied. For comparison of concentration scales, we mention that 1g./100 cc. of polysalt gives a solution whose stoichiometric bromide ion concentration is 0.04 normal. At the low concentration end of the curve, where simple electrolyte is in excess, the anomalous viscosity effect is very much reduced; at higher polymer concentrations, the effect of added salt becomes invisible. It is to be noted that the same ionic strength of different salts has about the same effect. Our interpretation is that the added simple salt furnishes an excess of *gegen*-ions and relatively fewer positive ions inside the chain are uncompensated with increasing added electrolyte; hence, the coil opens up less than in pure solvent, and the shear dependence decreases. (As will be seen shortly, the same conclusion regarding configuration can be drawn from the viscosity-concentration curves.) As the concentration of simple salt is increased, the flow dependence decreases; at 0.01 *N* potassium bromide, the effect is still visible, as can be seen by reference to TABLE 4, but at higher concentrations the change with shear was within the experimental error.

We now consider *the effect of added salt on the viscosity itself*. Curve 2 of FIGURE 1 gives the viscosity of the polybromide over the experimental range of concentration. It is concave-up over its entire course. Curve 3 shows the effect of the addition of 0.001082 *N* potassium bromide. The sharp climb at low concentrations is eliminated, and now a maximum appears in the curve. Curve 4 shows the effect of using 0.0335 *N* potassium bromide; here, the curve resembles that of an ordinary uncharged polymer, in that the curve is linear, with a small positive slope. The transition is shown more clearly in FIGURE 3, where the axis of ordinates has been magni-

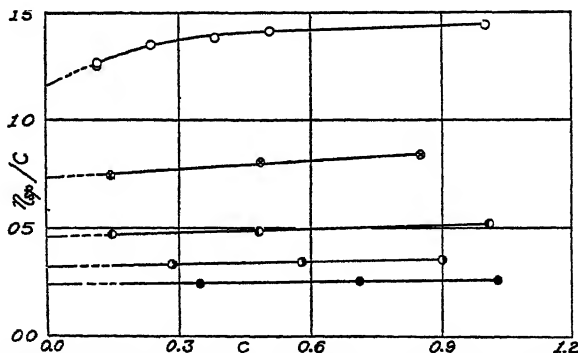


FIGURE 3. Viscosity at $\beta = 2,000$ of poly-salt in potassium bromide solutions of concentrations 0.01029 *N* (open circles), 0.0335 *N*, 0.1003 *N*, 0.2438 *N*, and 1.086 *N* (solid circles)

fied ten times. Similar results were obtained with potassium sulfate, magnesium sulfate, and magnesium bromide solutions as solvents. At low

concentrations of added electrolyte, the viscosity showed a maximum at the range of concentration where polyelectrolyte and added electrolyte were of the same order of magnitude of equivalent concentrations; with excess added electrolyte, the curvature disappeared and the curves became linear with a small positive slope. After the curves have become linear, the net effect of further addition of electrolyte is to depress the whole viscosity curve. These linear plots were extrapolated to zero polyelectrolyte concentration in order to obtain the figures of TABLE 12, where the limiting values are paired with

TABLE 12
LIMITING REDUCED VISCOSITIES IN THE PRESENCE OF ADDED ELECTROLYTE

KBr		MgBr ₂		K ₂ SO ₄		MgSO ₄	
μ	$[\eta]$	μ	$[\eta]$	μ	$[\eta]$	μ	$[\eta]$
0.01029	(1.16)	0.01222	(1.25)	0.00909	(0.75)	0.00712	(0.86)
0.0335	0.732	0.0519	0.713	0.08685	0.554	0.03375	0.704
0.1003	0.459	0.1575	0.433	0.2728	0.491	0.1075	0.590
0.2438	0.319	0.5541	0.288	1.060	0.472	0.3252	0.528
1.086	0.239					0.8140	0.504

the concentration of simple salt (given as ionic strength) in the corresponding solvent. The values in parentheses in TABLE 12 represent free-hand extrapolations of non-linear curves such as the top one of FIGURE 3. These limits represent the intrinsic viscosity of the polyelectrolyte *in an electrolytic environment*, it should be emphasized that they are, therefore, independent of polyelectrolyte concentration. They represent, thus, the most convenient figures on which to base a discussion of the effect of added electrolyte.

The first and most obvious property of the intrinsic viscosities is that they decrease with increasing concentration of added electrolyte. There is more involved, apparently, than a mere elimination of Coulomb repulsion by associative pairing off of charges in the polymer chain, because the intrinsic viscosities obtained by extrapolation in the presence of excess simple electrolyte are significantly *less* than that of the parent polymer. (If intramolecular repulsion were simply neutralized, the intrinsic viscosities should converge to about that of the parent polymer.) Internal attraction between ions of opposite charges within the polymer coil must be acting to produce the decrease in the size of the hydrodynamic unit, much like attraction between ions in a crystal acts to produce a structure of high density. We might make a very rough analogy to a small crystal of sodium chloride in which a fraction of the charges of the anions had been annihilated: there would still be ample potential energy to hold the unit together.

No theoretical basis is yet available for treatment of these data, aside from a feeling that the square root of ionic strength should be the controlling variable. A plot of the intercepts of TABLE 12 against reciprocal root of ionic strength shows some structure, so that the functional dependence is

more complicated than $[\eta] \sim \mu^{-1/2}$. At rather low concentrations of added salt, at least two properties of the polymer coil are changing with added ion concentration, namely, the average size of the coil itself and the mean range of its electrostatic interaction with the external electrolyte. It is possible that, with increasing concentration of external electrolyte, the first effect may reach a limit before infinite external ionic strength is reached, in that the polymer coil will reach its limit of compression or curling up at finite ionic strength. If this situation should obtain, the polyelectrolyte would begin to resemble, in its behavior, a rigid polyelectrolyte such as a protein molecule.

Oster¹⁵ has suggested that part of the ionic atmosphere of a colloidal ion is carried along with the latter, so that the equivalent hydrodynamic sphere has a radius equal to a constant plus $(1/\kappa)$ where κ is the Debye-Hückel reciprocal radius. Some screening certainly is present; just where to cut off the enhanced radius of the sphere is debatable. If we assume that the intrinsic viscosities of TABLE 12 are a measure of the volumes of the moving particles, and further assume that their radius is equal to R , their own radius, plus B/κ , a constant times the atmosphere radius reckoned on the external ion concentration, we would have

$$[\eta] = A(R + B/\kappa)^3 \quad (5)$$

whence a plot of the cube root of intrinsic viscosity against reciprocal square root of ionic strength should be linear. A test of these assumptions is shown in FIGURE 4 for the data on potassium and magnesium bromides and sulfates. Above an ionic strength of about 0.04, the curves are indeed linear, and furthermore 0.04 is the ionic strength above which the $\eta_{sp}/C - C$ curves of FIGURE 3 became linear.

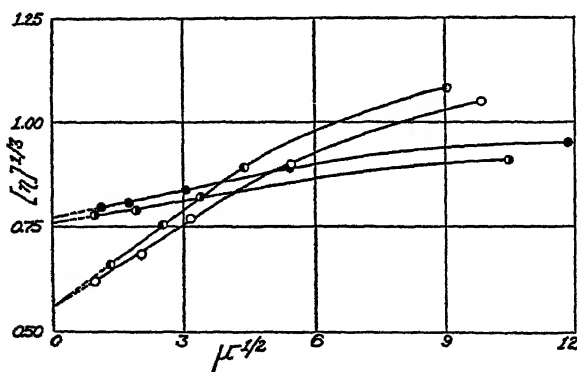


FIGURE 4. Dependence of extrapolated intrinsic viscosities on ionic strength of solvent. Left-shaded circles, magnesium bromide; open circles, potassium bromide; solid circles, magnesium sulfate; right-shaded circles, potassium sulfate.

Whether EQUATION 5 has real physical meaning is problematical at the moment, but the intercepts corresponding to infinite ionic strength probably

do. (It should be mentioned that the intercept at $\mu = \infty$ is not very sensitive to the negative exponent of μ chosen.) Both bromides extrapolate to $[\eta]_{\infty}^{1/3} = 0.556$, giving $[\eta]_{\infty} = 0.171$, which is to be compared with the value of 1.48 for the parent polymer. Obviously, the polyelectrolyte in the presence of an excess of added simple electrolyte is a much more compact structure than the neutral polyvinylpyridine from which it was prepared. It is significant too that the two sulfates extrapolate to a common value $[\eta]_{\infty} = 0.45$ and that this is about three times as large as the same limit for the bromides. Since the limit refers to a large excess of sulfate ions, all bromide ions inside the original polybromide coil must have been replaced by sulfate ions. Electropolar crossbonding¹⁸ by local interaction of sulfate ions with pyridonium groups pairwise could result in tangled structures which could not achieve the degree of compactness possible when the anion is only monovalent. It will be interesting to study the effects of other ions on these intercepts, and to attempt to evaluate them geometrically.

It may be dangerous to base a generalization on only two cations (potassium and magnesium), but it does seem reasonable to expect that the intercepts will be independent of the added cation and will depend only on the added counter ion. It should also be mentioned that the polybromide is precipitated by potassium ferro- and ferricyanide solutions.

The limiting slopes of the curves of FIGURE 4 should, according to EQUATION 5, be a measure of the size of the compacted polyelectrolyte. For the bromide, we obtain $R = 24$ A.U. if B is set equal to unity, as compared with 110 A.U. calculated from the intrinsic viscosity. To get agreement, B would have to be larger than four, which seems unreasonable; one would expect B values less than unity. While the assumptions, thus, obviously represent an over-simplification of the physical picture, we feel that the empirical form of EQUATION 5 may be useful in guiding theoretical work and that the limiting intercepts for infinite ionic strength will serve as a basis for comparing the properties of different polyelectrolytes.

We finally turn to a consideration of the *viscosity function of the polyelectrolyte in pure solvent*. Some earlier work¹⁰ in water, alcohol, and water-alcohol mixtures suggested that the function was of the form

$$\eta_{sp}/\sqrt{C} = A' + B'\sqrt{C} \quad (6)$$

but when the more dilute range was investigated in water, the $\eta_{sp}/\sqrt{C} - \sqrt{C}$ plots showed a curvature which became more pronounced the greater the dilution. A study of the data of TABLE 3 showed that they conformed to the empirical equation

$$\eta_{sp}/C = a/(1 + b\sqrt{C}) \quad (7)$$

as is shown by the test plot of FIGURE 5, where C/η_{sp} is plotted against the square root of polyelectrolyte concentration.

Obviously, we cannot have the function of EQUATION 6 representing data

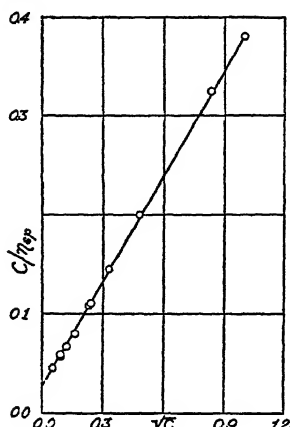


FIGURE 5. Test of EQUATION 2 for poly-salt in water at $\beta = 2,000$.

in alcohol and that of EQUATION 2 for data in water; both must be variants of the same function, with a parameter depending on dielectric constant causing the apparent shift. We are, therefore, led to propose the function

$$\eta_{sp}/C = \gamma + a/(1 + b\sqrt{C}) \quad (7)$$

as a more general formula. When $\beta\sqrt{C} \gg 1$, EQUATION 7 reduces to

$$\eta_{sp}/C = \gamma + a/b\sqrt{C}$$

which becomes EQUATION 6 on multiplying by the square root of concentration. When, on the other hand, $\gamma \ll \eta_{sp}/C$, EQUATION 7 reduces to EQUATION 2, the function of FIGURE 5.

Some experiments in nitromethane¹⁷ confirm the above conclusion. This solvent has a dielectric constant of 40, intermediate between that of water and ethanol. Here, neither EQUATION 2 nor EQUATION 6 fit the data, but the three-constant equation, EQUATION 7, gives linear test plots when $1/(\eta_{sp}/C - \gamma)$ is plotted against \sqrt{C} and the function reproduces the data within the limit of experimental error. A more general formula naturally calls for a linear term to be added to EQUATION 7, but we lack the data to substantiate the argument at the present time. Further work is in progress.

Summary

By addition of butyl bromide to polyvinylpyridine, a synthetic salt of high molecular weight (500,000) was obtained. Viscosities of this salt in water and in solutions of potassium and magnesium bromides and sulfates were measured at several rates of shear. Concentrations of polyelectrolyte covered the range from 0.003 to 1 g./100 cc. solution. In water, the specific viscosity increases sharply with dilution. Expansion of the polymeric coil

due to intramolecular Coulomb repulsion is suggested as the explanation for the rise in viscosity on dilution. In the presence of simple electrolytes, the increase is suppressed and linear viscosity curves of the conventional type appear. Their intercepts correspond to intrinsic viscosities which are much smaller than that of the uncharged parent polymer. For polyelectrolytes in the absence of other electrolytes, the empirical equation

$$\eta_{sp}/C = \gamma + a/(1 + b\sqrt{C})$$

represents the data so far obtained.

Bibliography

1. CRAGG, L. H. 1946. *J. Colloid Sci.* **1**: 261.
2. STAUDINGER, H. 1932. *Die hochmolekularen organischen Verbindungen*. Springer, Berlin.
3. EINSTEIN, A. 1906. *Ann. Phys.* **19**: 289; 1911; correction, *Ibid.*, **34**: 591.
4. EWART, R. H. 1946. *Advances in Colloid Science*. **II**: 197-297. Ed. H. MARK & G. S. WHITBY. Interscience, New York.
5. DEBYE, P. 1946. *J. Chem. Phys.* **14**: 636.
6. DEBYE, P. 1947. *Phys. Rev.* **71**: 486.
7. DEBYE, P. 1946. Letter quoted by L. H. CRAGG, *J. Colloid Sci.* **1**: 465.
8. Ref. 2 Part II D, by H. STAUDINGER & E. TROMMSDORFF.
9. KERN, W. 1938. *Z. phys. Chem.* **A181**: 249, 283.
10. FUOSS, R. M. & U. P. STRAUSS. 1948. *J. Polymer Sci.* **3**: 246.
11. BINGHAM, E. C. 1922. *Fluidity and Plasticity*, McGraw-Hill, New York.
12. CATHERS, G. I. 1948. Thesis. Yale University.
13. KROEPELIN, H. 1929. *Kolloid Z.* **47**: 294.
14. FLORY, P. J. 1945. *J. Chem. Phys.* **13**: 77.
15. OSTER, G. Private communication.
16. Electrolytes cross-bonded by primary valence can be prepared by adding ethylene dibromide to polyvinylpyridine. (Cf. GEORGE I. CATHERS. Thesis, Yale University, 1948.)
17. CATHERS, G. I. 1948. Thesis, Yale University.

June 22, 1949

ANTIBIOTICS DERIVED FROM
*BACILLUS POLYMYXA**

Consulting Editor: P. H. LONG

CONTENTS

	PAGE
Historical Aspects. By P. G. STANSLY	855
Isolation and Production of Polymyxin. By J. N. PORTER, R. BROSCARD, G. KRUPKA, P. LITTLE, AND J. S. ZELLAT	857
Effect of Various Factors on the Production of Polymyxin. By R. G. BENEDICT AND F. H. STODOLA	866
Antibiotics Derived from <i>Bacillus polymyxa</i> . By G. BROWNLEE	875
Comparative Biological Studies of Polymyxin and "Aerosporin." By H. J. WHITE, C. M. ALVERSON, M. J. BAKER, AND E. R. JACKSON	879
Comparative Biological Studies of Polymyxin A and Polymyxin D. By G. BROWNLEE, S. R. M. BUSHBY, AND E. I. SHORT	891
Chemical Studies on Polymyxin: Comparison with "Aerosporin." By P. H. BELL, J. F. BONE, J. P. ENGLISH, C. E. FELLOWS, K. S. HOWARD, M. M. ROGERS, R. G. SHEPHERD, R. WINTERBOTTOM, A. C. DORNBUSH, S. KUSHNER, AND Y. SUBBAROW	897
Chemical Evidence for the Multiplicity of the Antibiotics Produced by <i>Bacillus polymyxa</i> . By TUDOR S. G. JONES	909
The Chemistry of Polymyxin A. By J. R. CATCH, TUDOR S. G. JONES, AND S. WILKINSON	917
Infrared Studies. By R. C. GORE AND E. M. PETERSEN	924
Pharmacology of Polymyxin. By M. S. BRYER, E. B. SCHOENBACH, AND E. A. BLISS	935
<i>In Vitro</i> Studies of Polymyxin. By E. A. BLISS, C. A. CHANDLER, AND E. B. SCHOENBACH	944
The Pharmacology of Polymyxin A, B, and D. By G. BROWNLEE, S. R. M. BUSHBY, AND E. I. SHORT	952
Polymyxin: Assay Procedures. By E. T. REESE AND G. M. EISENBERG	968
A Simple Method for the Assay of Polymyxin in Blood and Urine. By P. G. STANSLY	980
The Effect of Polymyxin in The Treatment of Experimental Brucellosis AND Tularemia. By C. L. LARSON, B. N. CARLE, AND A. E. VERDER	982
The Clinical Use of Polymyxin. By E. B. SCHOENBACH, M. S. BRYER, AND P. H. LONG	987
Remarks on Clinical Results with Polymyxin A and B. By G. BROWNLEE	998

* This series of papers is the result of a Conference on Antibiotics Derived from *Bacillus polymyxa*, held by the Section of Biology of The New York Academy of Sciences, May 21 and 22, 1948. Publication made possible by grants-in-aid from Burroughs Wellcome & Co. (U.S.A.) Inc. and Lederle Laboratories Division, American Cyanamid Company and by the General Funds of the Academy.

Copyright, 1949, by

The New York Academy of Sciences

HISTORICAL ASPECTS

By P. G. STANSLY

Chemotherapy Division, Stamford Research Laboratories, American Cyanamid Company, Stamford, Connecticut

In highly active fields of research, it is, perhaps, not a rare occurrence for widely scattered groups of workers to be on the same trail of investigation unknown to each other. To examples of such incidents already known may be added that of the antibiotics of *Bacillus polymyxa*. In this instance, three laboratories independently and practically simultaneously reported on a new antibiotic substance or group of substances of unusual interest.

A paper entitled "Antibiotic Activity of *Bacillus polymyxa*," by R. G. Benedict and A. F. Langlykke of the Northern Regional Research Laboratories, was read before the annual meeting of the Society of American Bacteriologists on May 16, 1947. An abstract of this paper appeared in the July 1947 issue of the *Journal of Bacteriology*.¹ These authors noted the antibacterial activity of colonies or streaks of *Bacillus polymyxa* on agar plates and established that the active material was water-soluble and was readily produced free from the bacterial cells. A medium was described for obtaining culture filtrates active at a dilution of 1:1000 against *Brucella bronchiseptica*, and it was indicated that experiments were in progress to produce the substance on a large scale to determine its physical and chemical properties.

In July, 1947, there appeared in the Bulletin of the Johns Hopkins Hospital² a paper entitled "Polymyxin: A New Chemotherapeutic Agent" by P. G. Stansly, R. G. Shepherd, and H. J. White of the Stamford Research Laboratories of the American Cyanamid Company. This report was submitted for publication on May 10, 1947, and summarized investigations, begun in the summer of 1944, on a biologically inhomogeneous but highly purified antibiotic substance obtained from a soil isolate identified as *Bacillus polymyxa*, to which the authors gave the name "polymyxin." The production, isolation, and purification of the active principle was described and a microbiological assay method outlined. The physical, chemical, and, especially, the biological properties of purified material were discussed in some detail. Biologically, the antibiotic substance was characterized by its unique specificity for gram-negative bacteria and by the difficulty in obtaining resistant mutants from sensitive strains of bacteria. The material was shown to be a highly effective chemotherapeutic agent in several experimental infections produced by gram-negative bacteria in mice, and to have a favorable, acute, lethal to therapeutic dose. This paper was soon followed by others from the same laboratory in which the various aspects of the work on polymyxin were taken up at length.³⁻⁶

On August 23, 1947, there appeared in *Nature* a brief communication entitled "Aerosporin, An Antibiotic Produced by *Bacillus aerosporus*

Greer" by G. C. Ainsworth, A. M. Brown, and G. Brownlee of the Wellcome Physiological Research Laboratories in England.⁷ This work originated in February, 1946, and was submitted for publication on May 19, 1947. It was immediately apparent that aerosporin and polymyxin were either identical or closely related substances. Not only were the antibiotic-producing organisms similar (*B. aerosporus* is not a well known term and is considered by Bergey to be a minor synonym of *Bacillus polymyxa*⁸), but the indicated properties of the antibiotic were strikingly similar to those of polymyxin—notably in its selective action against gram-negative bacteria and the failure to obtain resistant strains from sensitive species. A later paper⁹ from the Wellcome Laboratories reported on the chemotherapeutic and pharmacological properties of aerosporin.

It seemed desirable to resolve promptly the question of the relationship of polymyxin and aerosporin. Arrangements were made, therefore, for the group at the Stamford Laboratories to exchange material with the Wellcome Laboratories in order that detailed chemical and biological studies, comparing the substances designated polymyxin and aerosporin, could be made on both sides of the Atlantic. The results of these investigations, among other things, are presented in certain of the papers which follow.

As will become evident, the main active components of polymyxin and aerosporin are not identical. They are, however, closely related chemically and practically indistinguishable biologically. Other active substances from *B. polymyxa* exist which appear to differ from the main constituents of polymyxin and aerosporin only in detail.^{10, 11} A common pattern of chemical constitution may, therefore, be anticipated in this group of antibiotics.

Bibliography

1. BENEDICT, R. G., & A. F. LANGLYKKE. 1947. Antibiotic activity of *Bacillus polymyxa*. *J. Bact.* 54: 24.
2. STANSLY, P. G., R. G. SHEPHERD, & H. J. WHITE. 1947. Polymyxin: A new chemotherapeutic agent. *Bull. Johns Hopkins Hosp.* 81: 43.
3. STANSLY, P. G., & M. E. SCHLOSSER. 1947. Studies on polymyxin: Isolation and identification of *Bacillus polymyxa* and differentiation of polymyxin from certain known antibiotics. *J. Bact.* 54: 549.
4. STANSLY, P. G., & M. E. SCHLOSSER. 1947. Studies on polymyxin: An agar diffusion method of assay. *J. Bact.* 54: 585.
5. STANSLY, P. G., M. E. SCHLOSSER, N. H. ANANENKO, & M. H. COOK. 1948. Studies on polymyxin: The production of fermentation liquor. *J. Bact.* 55: 573.
6. STANSLY, P. G., & N. H. ANANENKO. 1947. Resistance of polymyxin to some proteolytic enzymes. *Arch. Biochem.* 15: 473.
7. AINSWORTH, G. C., A. M. BROWN, & G. BROWNLEE. 1947. "Aerosporin", an antibiotic produced by *Bacillus aerosporus* Greer. *Nature* 160: 263.
8. *Bergey's Manual of Determinative Bacteriology*. 6th Ed. 1948. Williams & Wilkins. Baltimore.
9. BROWNLEE, G., & S. R. M. BUSHBY. 1948. Chemotherapy and pharmacology of aerosporin. *Lancet*, 254: 127.
10. BELL, P. H., J. F. BONE, J. P. ENGLISH, C. E. FELLOWS, K. S. HOWARD, M. M. ROGERS, R. G. SHEPHERD, R. WINTERBOTTOM, A. C. DORNBUSH, S. KUSHNER, & Y. SUBBAROW. 1948. Chemical studies on polymyxin: comparison with "aerosporin." *Ann. N. Y. Acad. Sci.* 51 (5): 897-908.
11. CATCH, J. R., T. S. G. JONES, & S. WILKINSON. 1948. The chemistry of polymyxin A. *Ann. N. Y. Acad. Sci.* 51 (5): 917-923.

ISOLATION AND PRODUCTION OF POLYMYXIN^{*}

By J. N. PORTER, R. BROSCARD, G. KRUPKA,
P. LITTLE, AND J. S. ZELLAT

Lederle Laboratories Division, American Cyanamid Company, Pearl River, N. Y.

Isolation of Bacillus polymyxa

In the course of routine examination of soil samples for the presence of antibiotic-producing organisms, a number of strains of the *Bacillus polymyxa* type were isolated. The possibility that some of these are *Bacillus macerans* cannot be entirely excluded, because differential tests were run on only a limited number of the isolates. However, all of them had been selected because they had produced zones of inhibition in plates seeded with *Mycobacterium tuberculosis* #607; had inhibited other organisms on non-seeded plates; or, in the case of three, had been contaminants in actinomycete cultures.

Bacillus polymyxa is a species which is relatively common in soils and other habitats. We have isolated 65 examples from soils of 22 states and believe that they represent a variety of physiologically differing strains. Apparently they differ in their ability to produce antibiotics on various media, amount of gas produced, mucin in pure culture, and in other ways. The isolate with which we have had the most experience is one which was designated B-71 by Dr. B. M. Duggar, who obtained it from a sample of Colorado soil. *In vivo* tests with many of the other isolates tended to show that none were better producers of polymyxin and several were distinctly inferior.

In Vitro Tests. We have utilized a bacterial spectrum of 15 representative pathogenic and non-pathogenic bacteria in testing broth filtrates of all newly isolated antibiotic-producing organisms. The agar streak-plate method of Waksman,¹ in which 1 cc. of the liquid to be tested is mixed with 9 cc. of agar and the solidified surface streaked with test organisms, was employed in each case. The results were in agreement with those reported by Stansly *et al.*² for polymyxin, and by Brownlee and Bushby³ for aeroporin, including the unique failure to inhibit *Serratia marcescens* and *Proteus vulgaris*. Fifty-six of the 64 isolates of *B. polymyxa* inhibited from one to ten of the test organisms included in the spectrum. Crude broths (as distinct from purified material) show a broader range of activity and, as did Stansly, we assume a second active material to be produced initially in small amounts.

Based on the ability of liquid filtrates to protect chicks against fowl cholera (caused by *Pasteurella multocida*), 18 of the earlier isolates of *B. polymyxa* were selected for comprehensive *in vivo* and *in vitro* testing to determine the more active. Roux bottles, shaker flasks, aerated 20-liter

^{*} The authors wish to acknowledge the assistance of Dr. L. M. Pruess, who carried out the tank fermentations, and of Mr. A. C. Dornbush, who performed the assays.

bottles, and 50-gallon aerated tanks were employed. All media consisted of 3 per cent corn steep liquor and from 1 to 3 per cent dextrose. 1 per cent dextrose was added aseptically after sterilization to the Roux bottles and shaker flasks. Three per cent dextrose was added before sterilization to the aerated bottles and 50-gallon tanks. The pH before sterilization in all cases was approximately 7.2. Streak-plate dilution assays, utilizing *Escherichia coli* A.T.C.C. 9637, *Shigella gallinarum* and *Pasteurella multocida*, were run on the broths. Fairly consistent differences in antibiotic activity were observed among the 18 isolates, and, in the majority of cases, the aerated tanks gave somewhat higher levels of potency.

TABLE 1
SUMMARY OF TESTS WITH FOWL CHOLERA
(2 cc. crude broth per injection)

Designation of isolate	Total No. of chicks tested	No. of survivals	Per cent saved
B-55	178	12	7
B-71	176	25	14
C-1	148	31	21
C-2	147	29	20
C-3	250	57	23
P-66	104	2	2
P-67	121	19	16
P-69	224	31	14
P-176	247	45	18
P-203	284	52	18
P-205	230	28	12
P-210	309	37	12
P-211	226	21	9
P-212	135	17	13
P-219	219	44	18
P-231	185	38	21
P-234	105	17	16
P-244	107	20	19
Controls	345	3	0.87

In Vivo Tests

FOWL CHOLERA (*Pasteurella multocida*). Screening of isolates was performed against *Pasteurella multocida* in chicks (because all controls die within two days), thus facilitating a larger number of tests with the physical facilities at hand. Two cc. of filtrate produced by one of the methods of fermentation listed above were injected into the peritoneum of a 3-day old chick. This was done $\frac{1}{2}$ hour after the injection of 0.5 cc. of a 10-3 dilution of a 6-hour broth culture of the pathogen. The tests were maintained for approximately 10 days after all controls were out, but the time varied occasionally, depending on whether the cages were needed for other tests.

TABLE 1 shows a summary of the results obtained with the 18 strains producing enough polymyxin to protect one or more of the chicks tested. Fourteen other strains did not protect on the initial tests. The remaining 32 isolates were not tested in this way. Of the 18 isolates affording some

protection 3 saved only from 2 to 9 per cent of the birds, while the remainder gave an average of about 18 per cent protection. The results agreed fairly closely with streak-plate assays on the same materials.

Grand totals of the results obtained with these 18 isolates with respect to the type of fermentation are given in TABLE 2. Again, the results agree with the streak-plate assays in that the tank fermentations yielded the highest results.

Ten cultures of *B. polymyxa* were obtained from the University of Wisconsin and were tested against fowl cholera in the manner outlined above.

TABLE 2
SUMMARY OF RESULTS WITH FOUR TYPES OF FERMENTATION

Type of fermentation	Total No. of chicks tested	No of survivals	Per cent saved
Roux bottles	1227	202	16.5
Shaker flasks	729	88	12.0
Aerated bottles	606	83	13.6
50 gal. tanks	826	163	19.7

Chicks inoculated with *Pasteurella multocida*

TABLE 3
SUMMARY OF 10 WISCONSIN STRAINS OF *B. polymyxa* TESTED AGAINST FOWL CHOLERA
(2 cc. crude broth per injection)

Wisconsin strain	No. of chicks tested	No. of survivals	Per cent saved
WC-6	50	0	0
WC-7	50	1	2
WC-8	50	0	0
W-9	50	7	14
W-12	50	1	2
W-19	50	6	12
W-57	50	11	22
W-362	50	2	4
W-419	50	5	10
W-510	50	4	8
Controls	30	0	0

The WC series (TABLE 3) yielded very little antibiotic, while the remaining 7 strains protected from 2 to 22 per cent of the chicks tested. Most of these seemed to be inferior to the strains already at hand.

In addition to testing crude broths, titration experiments were run with purified material derived from the B-71 strain. This was a dried eluate obtained by the streptomycin isolation procedure and assayed 800 units/mg. by the streak-plate method of assay, using *E. coli* as the test organism. As shown in TABLE 4, the number of survivals ranges from 30 per cent with 0.06 mg. (50 units) to 75 per cent with 2 mg. (1600 units). We did not go above 2 mg. per chick.

FOWL TYPHOID (*Shigella gallinarum*). The procedure in testing against fowl typhoid was varied somewhat from that given above. One cc. of a

10-3 dilution of 6-hour broth was administered almost simultaneously with 1 cc. of antibiotic by way of the peritoneum. Controls averaged between 50 and 100 hours for time of death and surviving birds were retained for three weeks. Substitution of 4 of the Wisconsin strains was made for the

TABLE 4
FOWL CHOLERA TESTS WITH PURIFIED POLYMYXIN

Amount injected	Total No of chicks tested	No of survivals	Per cent saved
2 mg	50	36	72
1 "	50	31	62
0.5 "	80	51	64
0.25 "	70	35	50
0.12 "	60	20	33
0.06 "	20	6	30

Material used was B-71 dried eluate assaying 800 *E. coli* streak plate units/mg. Chicks were 3 days old when injected and averaged 50 gm. in weight.

TABLE 5
SUMMARY OF TESTS WITH FOWL TYPHOID
(2 cc. crude broth per injection)

Designation of isolate	Total No of chicks tested	No of survivals	Per cent saved
P-67	167	10	6
P-69	172	4	2
P-176	190	5	2.5
P-203	194	4	2
P-205	185	4	2
P-210	184	9	5
P-211	185	12	6.5
P-212	175	12	7
P-219	159	4	2.5
P-231	164	3	2
P-234	180	23	13
P-244	171	7	4
C-1	187	16	8.5
C-2	203	15	7.5
C-3	200	24	12
B-71	173	25	14.5
W-9	46	1	2
W-19	86	2	2
W-57	68	2	3
W-419	68	3	4
Controls	154	0	0

two least active of the original 18 selected isolates. Survivals ranged from 2 to 14.5 per cent of the birds tested (TABLE 5).

HUMAN TYPHOID (*Eberthella typhosa*). None of the other isolates was deemed superior to the B-71 strain originally selected for chemical isolation purposes. Experiments were run against *E. typhosa* in mice utilizing B-71 dried eluate assaying 800 *E. coli* streak-plate units/mg. TABLE 6 shows the results of a typical experiment in which complete protection was ob-

obtained at the 0.5 mg. (400 units) level. The culture dose was 0.5 cc. of a 10-1 dilution of a 6-hour broth culture injected into the peritoneum. Treatment consisted of a single 0.5 cc. dose injected by way of the peritoneum immediately following injection of the pathogen.

Resistant Strains. A considerable delay in death among chicks infected with both *S. gallinarum* and *P. multocida* following injection of crude filtrates led us to surmise that some cells of the pathogenic population might be relatively resistant. A delay in death could then be due to the longer

TABLE 6
TEST WITH PURIFIED POLYMYXIN AGAINST *Eberthella Typhosa* IN MICE

Amount injected	Units injected	Mice protected	Mice dead	Average hours of death
1 mg	800	5	0	—
0.5 mg.	400	5	0	—
0.25 mg	200	4	1	168 0
0.12 mg	100	3	2	168 0
0.06 mg	50	3	2	168 0
0.03 mg	25	3	2	156 0
0.015 mg	12	1	4	156 0
Penicillin	250	1	4	36 0
Controls	0	0	5	17 2

Material used was B-71 dried eluate assaying 800 *E. coli* streak plate units/mg

TABLE 7
RESISTANT STRAINS OF *Shigella gallinarum*

Strain	Dilutions of B-71 eluate							
	1/100	1/200	1/400	1/600	1/800	1/1000	1/2000	1/4000
Original	—	—	—	—	—	—	—	—
18-B	+	+	+	+	+	+	+	+
18-E	—	—	—	—	—	—	—	—
18-F	+	+	+	+	+	+	+	+
19-D	+	+	+	+	+	+	+	+

— = Complete inhibition; + = no inhibition.

All strains but original were cultured from hearts of chicks which died after all controls were dead

incubation time needed by the limited number of cells unaffected by the antibiotic. Cultures were made from the hearts of several chicks originally inoculated with *S. gallinarum* and dying two or three days after all controls were out. Such cultures proved quite infective for the chicks upon retest, and, when tested against a strong polymyxin solution in streak plates, three of them showed a great contrast from the original culture (TABLE 7).

A program was initiated to select resistant strains by *in vitro* methods in order to check the above results. Using an active B-71 eluate at dilutions up to 1 part in 16,000, streaks were made from 24-hour broth cultures of 6 pathogens. A broth culture was made from a colony of each pathogen at

the lowest dilution showing any growth. The process was repeated for a period of several transfers. From the limited data at hand (TABLE 8), it appeared that resistance was developed to polymyxin by *S. gallinarum* and *P. multocida*, but not by *E. coli*, *E. typhosa*, *K. pneumoniae*, and *S. pullorum*.

Fermentation. All the material which we have employed for purification purposes has been produced from the B-71 strain in aerated tanks. The first runs were made in 200-gallon tanks, but we are currently using larger fermentors. The data given below apply primarily to the 200-gallon tanks.

Medium—	3 per cent corn steep liquor 3 per cent dextrose
Aeration—	0.7 liters of air per liter of mash per minute.
Stirring—	135 r.p.m.
Temperature—	26–28°C.
pH—	Initially 7.2; after sterilization 6.5; 42 hrs. 6.0; 64 hrs. 6.7.
Yields—	Approximately 400 streak-plate units per cc. at 66 hrs. when the tanks are harvested.

Repeated runs on other types of media have yielded little or no potency on this medium with the B-71 strain. Conversely, a number of trials have

TABLE 8
In Vitro SELECTION OF RESISTANT STRAINS

Pathogen	Initial complete inhibition	Complete inhibition after 7 transfers
<i>Shigella gallinarum</i>	1/2000	1/50
<i>Pasteurella multocida</i>	1/2000	1/50
<i>Escherichia coli</i>	1/4000	1/2000
<i>Eberithella typhosa</i>	1/4000	1/4000
<i>Klebsiella pneumoniae</i>	1/8000	1/4000
<i>Salmonella pullorum</i>	1/4000	1/8000

B-71 eluate was the material used in these tests.

shown that the Stamford strain produces but slight results on the corn steep liquor medium. The two strains appear to differ further in their sensitivity to increased aeration and agitation, in the amount of gas produced, and in the amount of mucin formed.

Purification and Isolation of Polymyxin

The isolation of polymyxin from media fermented in 20-liter aerated bottles was accomplished by using the charcoal adsorption method developed by Carter *et al.*⁴ for the purification of streptomycin. The eluates referred to in the first part of this report were obtained in this way. The liquor produced in deep tank fermentation presented a serious problem of filtration in that a mucoid material was produced which made filtration impossible. This problem was solved by shifting to a method which has been employed successfully by us in the treatment of other polypeptides, namely, salting the active principle into an organic solvent.

Salting into Isopropanol. To the liquor produced *via* deep-tank fermentation is added a one-quarter volume of isopropanol and stirred for one half-hour. To the alcoholic solution is added enough $(\text{NH}_4)_2\text{SO}_4$ to cause the two fractions to separate. The alcohol is decanted and another one-quarter volume of isopropanol added and stirred for an additional one-half hour. The two decanted aliquots are pooled and filtered. At this point, the activity may be either adsorbed directly from the isopropanol, or the latter may be distilled *in vacuo* at a temperature no greater than 50–60°C. If distillation is employed, enough water is added toward the end of the process to make a light syrup. This is then filtered, placed in bottles, and frozen and dried *in vacuo* for future use.

Aluminum Oxide Adsorption. For the adsorption of polymyxin from the isopropanol extract or from a methanolic solution of the dried distillate, a glass column packed with 40–80 mesh aluminum oxide is employed. Before adsorption, the column is washed with 10 per cent H_2SO_4 , followed by thorough washes with distilled water until the pH of the filtrate is 3–4. The column is then dried by passing methanol through it. The active isopropanol or methanol fraction is then passed through the column by gravity. Following adsorption the column is washed with water-methanol mixture which removes most of the color and solids with a loss of 10–15 per cent of the activity. Elution is then accomplished with successive volumes of distilled water.

Crystallization of Free Base. The following is a typical experiment resulting in crystallization of polymyxin: 200 gm. of dried isopropanol extract were stirred with 2.8 liters of methanol and filtered. The filtrate contained 193 gm. of solids and a total of 32,200 *Shigella* units of activity.* The adsorption column was a 3 inch pyrex glass tube filled to a depth of 26 inches with 40–80 mesh aluminum oxide. The column was washed with one liter of 10 per cent H_2SO_4 , followed by washes with distilled water and drying with methanol. As the alcoholic extract passed through the column by gravity, most of the black-red color was adsorbed. Most of this was washed out with 4 liters of 80 per cent methanol followed by 4 liters of 50 per cent methanol. Water elution followed. Results are given in TABLE 9.

The second and third eluates were adjusted to pH 10 with N NaOH and filtered. The filtrate was adjusted to pH 7 with N HCl, and then 9 grams of picric acid dissolved in hot water were added. After stirring, the solution was allowed to stand overnight at 4°C. The supernatant was then decanted and the picrate, which adhered to the beaker, was dissolved in one liter of acetone and filtered by gravity. Acetone saturated with HCl gas was added drop by drop to the acetone. When no more hydrochloride could be precipitated, the acetone was placed in the chill room for several hours. The hydrochloride was then filtered off by suction and dried *in vacuo*.

* Because the streak-plate method is too crude for analyzing purification procedures, a cup-plate zone assay was devised using *S. gallinarum* as the test organism. One *Shigella* unit was roughly equivalent to 200 streak-plate *E. coli* units.

Three grams of this hydrochloride (approx. 4-5 *Shigella* units/mg.) were dissolved in 50 cc. of distilled water. The water was cooled to freezing and ammonia gas introduced with a pipette. The liquid solidified and was

TABLE 9
ADSORPTION AND ELUTION OF POLYMYXIN BY COLUMN

Steps in the process	Volume	Total solids	Total activity in <i>Shigella</i> units
Alcohol extract	2.8 liters	193.0 gm.	32,000 units
Adsorption filtrate and 80% MeOH wash	7.0 liters	170.6 gm.	inactive
50% MeOH wash	3.77 liters	5.714 gm.	inactive
1st H ₂ O elution	1.02 liters	1.632 gm.	816 units
2nd H ₂ O elution	1.01 liters	6.464 gm.	18,937 units
3rd H ₂ O elution	1.02 liters	2.555 gm.	3,577 units
4th H ₂ O elution	1.95 liters	2.493 gm.	448 units
Total eluted with water	5.0 liters	13.144 gm.	23,778 units

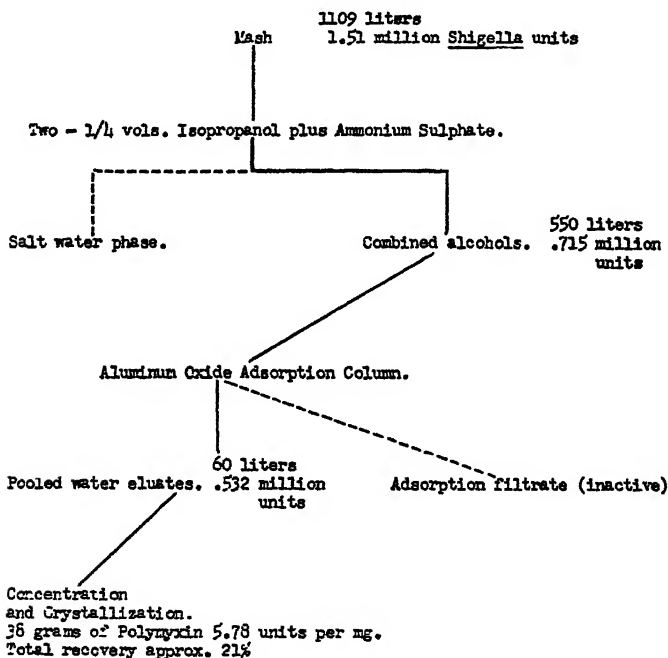


FIGURE 1. Purification of polymyxin, flow sheet.

placed in the chill room to complete the crystallization. After two hours, 50 cc. of cold 28 per cent NH₃ were added and the mass stirred, then filtered by suction. The crystals were washed with cold water and dried *in vacuo*.

When observed with an analyzing microscope, the crystals showed birefringence, and had no definite structure. They gave a strong violet

coloration with $\text{CuSO}_4\text{-NaOH}$ and had a point of decomposition between 220–250°C. They assayed approximately 10 *Shigella* units, mg.

Production of Polymyxin. Essential features of a typical run are given in the flow sheet (FIGURE 1). The column used was 8 inches in diameter by 36 inches in length. It was prepared for adsorption by the method given above. The 550 liters of isopropanol extract passed through the column were followed by 50 liters of methanol and then 75 liters of 50 per cent water-methanol. Seventy-five liters of distilled water were used for elution. The eluate was concentrated to 4 or 5 liters, filtered, cooled to the freezing point, saturated with ammonia, and allowed to stand two hours. The crystals were filtered off by suction, washed with cold water, and dried *in vacuo*. The recovery in the crystals was approximately 21 per cent of the activity initially present in the harvested liquid.

Summary

Sixty-five isolates of *Bacillus polymyxa* were obtained from fifty-two sources, mostly soil samples. There seemed to be important differences among these isolates with respect to their antibiotic production, as well as certain other factors. Ten cultures obtained from Wisconsin showed similar differences. The isolates demonstrated varying degrees of effectiveness against *Shigella gallinarum*, *Pasteurella multocida*, and *Eberthella typhosa* as determined by *in vivo* experiments. There is some evidence that a few bacteria develop resistance to polymyxin, although this does not seem to be true of the majority.

Polymyxin may be produced in aerated tanks after a fermentation period of two or three days at a temperature of 26–28°C. The active principle can be isolated in two ways: (1) a charcoal adsorption, followed by an acid-methanol elution; or (2) salting into isopropanol, followed by adsorption on an aluminum oxide column using water as an eluant. Concentrated eluates contain 30–40 per cent of the initial activity. The combined eluates are then purified by making the picrate decomposition in acetone. Ammonia is used to complete the crystallization of the antibiotic.

Bibliography

1. WAKSMAN, S. A., & H. C. REILLY. 1945. Agar-streak method for assaying antibiotic substances. *Ind. Eng. Chem. Anal. Ed.* 17: 556.
2. STANSLY, P. G., R. G. SHEPHERD, & H. J. WHITE. 1947. Polymyxin: A new chemotherapeutic agent. *Bull. Johns Hopkins Hosp.* 81: 43.
3. BROWNLEE, G., & S. R. M. BUSEBY. 1948. Chemotherapy and pharmacology of aerosporin. *Lancet.* 1: 127.
4. CARTER, H. E., R. K. CLARK, S. R. DICKMAN, Y. H. LOO, P. S. SKELL, & W. A. STRONG. 1945. Isolation and purification of streptomycin. *J. Biol. Chem.* 160: 337.

EFFECT OF VARIOUS FACTORS ON THE PRODUCTION OF POLYMYXIN

By R. G. BENEDICT AND F. H. STODOLA

Fermentation Division, Northern Regional Research Laboratory, Peoria, Illinois*

Introduction

In May 1947, Benedict and Langlykke¹ reported the production of an antibiotic from strains of *Bacillus polymyxa*. It developed at that time that Stansly, Shepherd, and White,² of the American Cyanamid Company, had independently discovered the same antibiotic and had, as a matter of fact, conducted considerable research on phases of production, *in vitro* testing, etc. They designated the antibiotic "polymyxin." In August of last year, Ainsworth, *et al.*,³ working at the Wellcome Laboratories in England, also independently reported the production of an antibiotic, designated "Aerosporin," from *Bacillus aerosporus*, which most authorities believe is synonymous with *B. polymyxa*.

During the period since our previous report, we have exchanged information, cultures, and samples of polymyxin with the Cyanamid group and we welcome the opportunity of similar cooperation with the Wellcome Laboratory group and other workers.

As stated in our report a year ago, we were searching for antagonists which would be effective against the *Brucella* group and other gram-negative organisms. Our preliminary studies showed that certain soil isolates of *Bacillus polymyxa* were capable of producing antibiotic substances of considerable potency. Taking advantage of our collection of this organism, which had been enlarged for use in the production of 2:3 butylene glycol, we made observations on the antibacterial spectra of 32 known strains as well as 7 new isolates. We also found that it was possible to concentrate the antibiotic activity obtained in the filtrate from a better yielding strain. The present report is based particularly upon work which has been conducted at the Northern Regional Research Laboratory during the past year.

In our experience, strains of *Bacillus polymyxa* were found to produce, on agar plate media, an apparently water-soluble factor which markedly inhibited the growth of some Gram-positive types. This factor was studied by Stansly and Schlosser,³ who extracted it from the cells of *B. polymyxa* as a water-soluble, ethanol-soluble substance which was highly active against *Staphylococcus aureus* and inactive against *Escherichia coli*. We have found no adequate explanation for the apparent loss in water-solubility which this factor undergoes during the process of partial purification.

Preliminary Studies on Various Media for the Production of Polymyxin. *Bacillus polymyxa* is not a fastidious organism. Katznelson and Lochhead⁵

* One of the laboratories of the Bureau of Agricultural and Industrial Chemistry, Agricultural Research Administration, U. S. Department of Agriculture.

found that growth factor requirements in a synthetic medium were satisfied by the addition of 0.1 μ g. biotin per liter. Thiamin was not essential, but low levels added to the medium caused growth stimulation of some strains and inhibition of others. As a source of organic nitrogen, yeast extract was superior to tryptone or peptone.

Our preliminary studies soon indicated that good growth of the organism is not always correlated with high yields of polymyxin. Furthermore, yields varied greatly with different substrata. In our preliminary tests, active culture filtrates were obtained from only 2 of 10 different media tested. The routine procedure of sterilizing crude beers by sintered-glass filtration proved fortunate, because we later found that polymyxin from crude culture liquor is highly adsorbed on Seitz filter pads.

Of the two favorable media mentioned above, a corn steep-glucose-calcium carbonate preparation gave the most active filtrates. The relatively low yields of polymyxin obtained in this medium, when compared with later work, were due partly to the conditions of fermentation. Its use was discontinued when trials with the yeast extract-mineral salts-dextrose medium of Stansly, Shepherd, and White² was found to produce significantly higher yields. As indicated by these workers, initial pH adjustment, addition of separately sterilized dextrose, use of small inocula, and fermentation at 25°C. are important factors in attaining good yields of polymyxin in shallow surface cultures or in deep fermentation vessels with moderate aeration.

Polymyxin Assay Methods. Different methods of assaying polymyxin were investigated. Our first assays were done by a serial dilution method using *Brucella bronchiseptica* NRRL B-140 as the test organism. This procedure proved unusually tedious since all culture liquors had to be filter-sterilized prior to test. In an attempt to develop a more rapid assay, *Escherichia coli* was investigated in connection with a 4-hour turbidimetric method. Reproducible results were difficult to obtain and this latter method was abandoned. In the meantime, Stansly and Schlosser⁶ developed a paper disc-agar diffusion method of assay with *E. coli* (MacLeod strain) as the test organism. Since the antibiotic polymyxin diffuses very slowly through agar, and since *E. coli* multiplies very rapidly, the use of this organism required that the assay plates be incubated at 25°C. for 18 hours, followed by 37°C. for 6 hours. *Brucella bronchiseptica* NRRL B-140 was found to be equally sensitive to polymyxin, and Benedict and Stodola⁷ found that its use in a paper disc method offered certain marked advantages. The zones of inhibition of both standard preparations and crude fermentation liquors had sharply defined edges, and assay plates could be incubated directly at 37°C. and read after 14 to 16 hours.

A polymyxin assay plate, with *Brucella bronchiseptica* B-140 as the test organism, is shown in FIGURE 1. The assay strain is remarkably stable and has been carried on artificial culture media for 10 years. Since strains of the organism rarely infect humans, the culture may be safely handled with the usual precautions afforded potential pathogens. Our plate unit of polymyxin is the same as that of Stansly and Schlosser.⁶ A partially purified

polymyxin hydrochloride standard and a subculture of the Cyanamid strain of *B. polymyxa* designated NRRL B-719) used to produce it, were kindly supplied by the Cyanamid group. All subsequent polymyxin standards were evaluated against this standard.

Ultraviolet Irradiation of Bacillus polymyxa Since ultraviolet irradiation proved so effective in producing higher penicillin-yielding mutants of *Penicillium chrysogenum* as shown by Backus *et al.*,⁵ this technique was tried on *Bacillus polymyxa*. The strains irradiated were NRRL B-719, B-

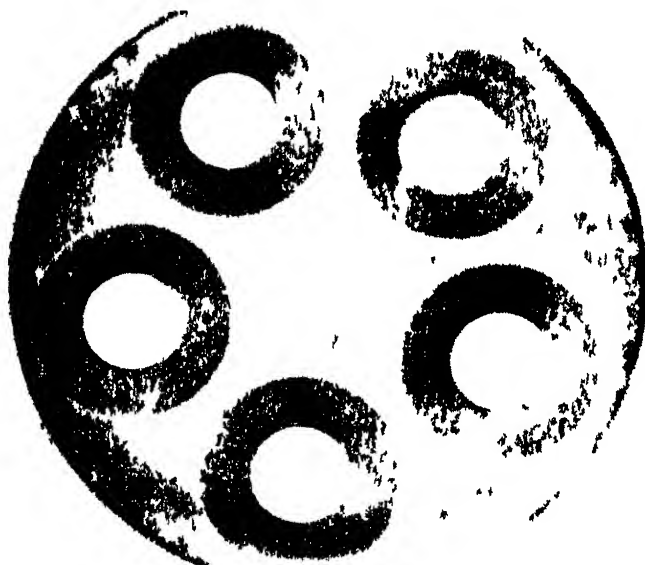


FIGURE 1. Polymyxin assay plate with *Bacillus polymyxa* NRRL B-719. Each paper disc was saturated in a solution containing 1.2×10^6 units polymyxin per ml.

719 HS, the same, but with spores heat-shocked for 1 minute at 90°C, and Adams and Ledingham's strain No. C-40(2). In brief, the technique was as follows: One loopful of a 24-hour broth culture was streaked across the center of a plate containing a soluble starch-peptone-yeast extract medium with 3 per cent agar. A sterile Pyrex glass "T" was used to spread the inoculum in a radial pattern on the agar plate. A set of plate cultures from each test strain was exposed to ultraviolet irradiation with covers removed, for intervals of 0, 1, 2, 4, 7, and 10 minutes. The apparatus employed in these studies was that used by Raper and Fennell⁹ in the production of higher yielding penicillin X mutants from *Penicillium chrysogenum* NRRL 1984. It consisted of a 4-watt General Electric germicidal lamp of

bent tube construction emitting 500 milliwatts of 2,537 Å (95 per cent of the total irradiation) as the source of ultraviolet light. The lamp was supported on a ringstand in a horizontal position 11.4 cm. above the base.

After 3 days' incubation at 25°C, the plates were examined for survivors. TABLE 1 shows the approximate numbers of surviving cells which developed colonies. The treatment appears to be quite drastic, as indicated by the reduction of viable cell numbers. Assuming that approximately equal numbers of *B. polymyxa* cells were streaked on all plates, this table shows that there is considerable variation between strains in their ability to withstand ultraviolet treatment.

Testing of Ultraviolet Mutants. Production of polymyxin in submerged culture appeared to be more practical than still culture production in shallow layers of medium. Laboratory tests, therefore, were conducted in shaken-flask cultures. The selected yeast extract-mineral salts-glucose medium was employed in 150-ml. quantities in 1-liter Erlenmeyer flasks, and each

TABLE 1
COLONY FORMATION IN STRAINS OF *Bacillus polymyxa* AFTER EXPOSURE TO ULTRAVIOLET IRRADIATION

Irradiation time (in minutes)	B-719	B-719 HS	C-40 (2)
0	—*	—	—
1	130	21	500-1000†
2	120	4	500†
4	45	3	300
7	5	0	90
10	1	0	55

* Colonies too numerous to count

† Approximate count

was inoculated with a single colony from a plate. The flasks were agitated on a rocker shaker having 88 oscillations per minute and a 3-inch stroke. Incubation was at 25°C. Samples were taken for pH and assay on the third, fourth, and fifth days. In a good fermentation, the pH drops from 7.5 at 0 hours to about 6.2 at 72 hours and rises to 6.8 at 96 hours. However, in the first set of shaken-flask tests, the pH of the B-719 unirradiated control, and also practically all of the ultraviolet-treated isolates of this strain had dropped to 4.7 and no polymyxin was produced. The beneficial effect of adding sterile calcium carbonate and sterile dextrose to some of these "sour" flasks is shown in TABLE 2. The results show that some polymyxin was produced after the neutralization of the "sour" culture liquors with calcium carbonate. More significant, however, was the observation that polymyxin production was greatly stimulated in the neutralized beers, to which additional sterile dextrose solution had been added. If allowed to incubate for 2 days at 25°C., still culture flasks, with or without calcium carbonate, could then be placed on the shaker, and significantly higher yields of polymyxin were produced at 4 days than in comparable still culture

flasks of the same age. However, if sterile calcium carbonate was added prior to inoculation with B-719 and the flasks shaken at 25°C., little or no polymyxin was produced even though the problem of souring of the medium had been eliminated.

It was obvious that a change either in the medium or in the method of fermentation should be made before testing further the production of polymyxin by NRRL B-719 and some of its ultraviolet-induced mutants. Stansly and co-workers¹⁰ had found that, of a number of cheaper protein sources, soybean oil meal was the only substitute for yeast extract which gave equivalent polymyxin production. In our laboratory, excellent results were obtained by substituting 0.5 per cent "Nutrisoy XXX," a solvent-extracted soybean meal, in place of yeast extract. The tests were run in Erlenmeyer flasks, with each containing 145 ml. of soya meal-mineral salt

TABLE 2
RECOVERY OF POLYMYXIN PRODUCTION IN "SOUP" FLASKS BY THE ADDITION OF CALCIUM CARBONATE AND DEXTROSE

Strain No.	+CaCO ₃ at 72 hours+		Dextrose at 116+ hours+	
	72 hr. assays		130 hr. assays	
	pH	u/ml. assay	pH	u/ml. assay
B-719 unirradiated	4.7	0	6.5	0
UV-24	4.8	0	6.3	28
UV-30	4.7	0	6.6	84
UV-35	4.7	0	6.6	42
UV-39	4.8	0	6.5	8
			6.4	270
			6.5	120
			6.7	96
			6.8	105

* Ultraviolet-treated isolates from B-719.

base adjusted to pH 7.7 before sterilization. Before inoculation, 5 ml. sterile 30-per cent dextrose solution and 1.5 grams of dry sterile calcium carbonate were added. The flasks were agitated at 25°C. on a rocker-type shaker. Souring of the medium did not occur in the absence of calcium carbonate, but adding it to the medium resulted in higher yields of polymyxin.

Polymyxin production by the stock strain of NRRL B-719 and 74 colonies developing from ultraviolet irradiated cells of the same were studied, along with the effect of feeding different amounts of sterile dextrose solution to the flasks at various times during the fermentation. Colonies from irradiated B-719 HS (heat shocked) and from irradiated C-40(2) produced no more polymyxin than the unirradiated controls. A number of ultraviolet-induced mutants from B-719 (not heat shocked) gave considerably higher yields than the unirradiated parent. Many combinations were tried in the process of feeding additional dextrose, but 1 per cent added between the second and third days of fermentation was the only one which con-

sistently doubled the polymyxin yields. Some strains of *B. polymyxa* produce large amounts of gummy material when the carbohydrate content of the medium is increased. A total of 2 per cent dextrose, half at the start of fermentation and the remainder added at 48 to 72 hours, did not cause excessive slime formation by any of the test strains. TABLE 3 shows the combined effects of ultraviolet irradiation and feeding of sterile dextrose on *B. polymyxa*, NRRL B-719, and three selected ultraviolet mutants of the same. At 112 hours, the polymyxin titers in the flasks receiving additional dextrose were twice those of the unfed controls. Strain UV-37 appeared considerably better than the unirradiated parent. Colony isolates were obtained by plating out all ultraviolet-treated strains which yielded 400 units per ml. or more. Four isolates from UV-37 produced 900 to 1000 units per ml. In subsequent experiments, these strains consistently pro-

TABLE 3
EFFECT OF ULTRAVIOLET IRRADIATION AND FEEDING OF STERILE DEXTROSE ON POLYMYXIN PRODUCTION BY *B. polymyxa* NRRL B-719

Strain	Dextrose fed	Age in hours at assay, plate units polymyxin per ml.			
		64	88	112	136
B-719 (control)	None	199	184	132	—
	1% at 64 hrs.	192	264	256	232
UV-24 [*]	None	212	232	164	—
	1% at 64 hrs.	204	232	352	212
UV-33 [*]	None	324	194	124	—
	1% at 64 hrs.	256	240	256	184
UV-37 [*]	None	160	220	176	—
	1% at 64 hrs.	180	500	360	184

* Ultraviolet-treated mutants from B-719.

duced from 500 to 900 units per ml. in shaken-flask cultures. Since these yields are so much greater than those from the unirradiated parent, we assume that ultraviolet treatment has produced relatively stable mutants and is a procedure which others might profitably use to increase polymyxin production in their own strains. Progeny of these strains were irradiated, but no higher yielding mutants resulted from these "secondary" exposures.

In shaken flasks with the soya meal-calcium carbonate medium there is some relationship in color between the supernatant fluid and polymyxin yields. Upon standing, a dark brown color was noted in the high-titer flasks as compared with a light straw yellow for the low-titer flasks. An attempt was made to correlate color intensity with polymyxin yields, using a spectrophotometer. However, the correlation was not sufficiently significant to allow use of color measurement as an analytical procedure for polymyxin assay.

Polymyxin Production by Strains other than B-719. Emphasis has been placed on polymyxin production by B-719 and mutants derived from it. In addition to these, 24 strains of *Bacillus polymyxa* from our culture collection and 11 strains isolated from soils were tested in the soya medium by the shaken-flask method. Compared with strain B-719 (unirradiated) averaging 200 units polymyxin per ml., 10 were better, 11 equal, and 14 produced less than 200 units per ml. The best strain of this group was a soil isolate, NRRL B-694, which without irradiation has consistently produced from 600 to 800 units per ml. with a maximum yield of 1290 units per ml.

In an effort to increase our yields of polymyxin, additional variations in the constituents of the medium were tried. Twenty of the above 35 strains produced significantly more polymyxin when hexane-extracted peanut meal was substituted for the same amount of soya meal. Of various concentrations tested, peanut meal at a level of 1.5 per cent appeared un-

TABLE 4
A COMPARISON OF POLYMYXIN PLATE UNITS AND SERIAL DILUTION ACTIVITY OBTAINED WITH FOUR STRAINS OF *B. polymyxa*

Strain No.	Plate units/ml.*	Dilution units/ml.*
NRRL B-503	262	36,000†
B-367	174	9,000
B-719-UV-37	710	15,000
B-694	665	9,000

* Assays at 93 hours against *Brucella bronchiseptica* NRRL B-140.

† Ratio of plate to dilution units is approximately 1:10.

usually favorable. As an inorganic source of nitrogen, 2 per cent ammonium citrate or di-ammonium hydrogen phosphate was equally as good as the 2 per cent ammonium sulfate normally used in the medium. Of the carbohydrate materials tested in polymyxin production by Stansly and co-workers and by ourselves, dextrose and sucrose were superior to lactose or soluble starch. Commercial hydrol, a residual molasses obtained in the manufacture of corn sugar, is rich in glucose and gentiobiose. We found that, when used on a comparable glucose basis, it proved a better carbohydrate source for polymyxin production by some strains than dextrose alone.

Formation of Different Polymyxins by Some Strains of B. Polymyxa. It was suggested to us by Reese and Eisenberg¹¹ that some strains of *B. polymyxa* produce polymyxin or polymyxins which do not readily diffuse through agar. A relatively simple procedure to show that some strains produce types which are less readily diffusible than the polymyxin produced by B-719-UV-37 and B-694 may be seen in TABLE 4. Four strains of *B. polymyxa* were grown in the same medium and 93 hour values on the crude beers determined by our paper disc-plate method and by a serial dilution method, with *Brucella*

bronchiseptica B-140 as the test organism. Although the plate units per ml. in the crude beers from strains B-719-UV-37 and B-694 were approximately 3 times greater than those from B-503, the dilution units from this culture were 2.4 times those from B-719-UV-37 and 4 times those from B-694. Strain B-367, showing one-fourth the plate unit activity of B-694, was equally active by the serial dilution method. These results suggest that B-503 and B-367 produce different polymyxins than those formed by B-693 and B-719-UV-37.

Polymyxin Production on a Semi Pilot-Plant Scale. A number of 15-liter fermentations were run in 22-liter round-bottom flasks. The aeration rate was one-fifteenth volume of sterile air per volume of culture liquor per minute. Under these conditions, 400 units per ml. was the highest yield of polymyxin attained. In the preparation of polymyxin concentrates from crude beers of equal unitage, higher potency material, on a weight basis, was obtained from the yeast extract medium than from the soya or peanut meal preparations. The carbon adsorption treatment, which removed all of the polymyxin activity from the clarified yeast extract beer, removed only half of the activity from culture filtrates of soya or peanut meal medium. The final product obtained from the purification of these latter filtrates contained more inactive protein than the product derived from the yeast extract filtrates.

Summary and Conclusions

Polymyxin production by 35 strains of *Bacillus polymyxa* has been studied in various media under different conditions. Ultraviolet radiation produced higher yielding polymyxin mutants from one of four strains irradiated. The feeding of sterile dextrose during the fermentation to all 35 test strains greatly increased the polymyxin yields. Soya and peanut meals were superior to yeast extract as protein sources in the medium, and represent cheaper and more available substitutes for large-scale production. Evidence for the formation of different polymyxins by some strains of *B. polymyxa* is based on slower rates of diffusion and marked differences in activities when compared by two methods of assay.

Bibliography

1. BENEDICT, R. G. & A. F. LANGLYKKE. 1947. The antibiotic activity of *Bacillus polymyxa*. J. Bact. 54: 24-25.
2. STANSLEY, P. G., R. G. SHEPHERD, & H. J. WHITE. 1947. Polymyxin: A new chemotherapeutic agent. Bull. Johns Hopkins Hosp. 81: 43-55.
3. AINSWORTH, G. C., A. M. BROWN, & G. BROWNLEE. 1947. 'Aerosporin,' an antibiotic produced by *Bacillus aerosporus* Greer. Nature 160: 263-264.
4. STANSLEY, P. G. & M. E. SCHLOSSER. 1947. Studies on polymyxin: Isolation and identification of *B. polymyxa* from certain known antibiotics. J. Bact. 54: 549-556.
5. KATZNELSON, H. & A. G. LOCHHEAD. 1944. Studies with *B. polymyxa*. III. Nutritional requirements. Can. J. Research 22C: 273-279.
6. STANSLEY, P. G. & M. E. SCHLOSSER. 1947. Studies on polymyxin: An agar diffusion method of assay. J. Bact. 54: 585-597.
7. BENEDICT, R. G. & F. H. STODOLA. 1948. Modification of an agar diffusion method of assay for polymyxin. J. Bact. 55: 286.

8. BACKUS, M. P., J. F. STAUFFER, & M. J. JOHNSON. 1946. Penicillin yields from new mold strains. *J. Am. Chem. Soc.* **68**: 152-153.
9. RAPER, K. B. & D. I. FENNELL. 1946. The production of penicillin X in submerged culture. *J. Bact.* **52**: 761-777.
10. STANSLY, P. G., M. E. SCHLOSSER, N. H. ANANENKO, & M. H. COOK. 1948. Studies on polynyxin: Production of active fermentation liquor. *J. Bact.* **55**: 573-578.
11. REESE, E. T. & G. M. EISENBERG. 1947. Personal communication.

ANTIBIOTICS DERIVED FROM *BACILLUS POLYMYXA*

By GEORGE BROWNLEE

The Wellcome Physiological Research Laboratories, Beckenham, Kent, England

The antibiotic at first referred to as "Aerosporin" and now known as Polymyxin A was discovered at The Wellcome Physiological Research Laboratories, in the course of routine screening of possible antibiotics in 1945.

The productive strain of the organism was first identified as *Bacillus aerosporus* Greer, an organism which is now recognized to be identical with *B. polymyxa*, a widely distributed bacterium whose natural habitat appears to be the soil.

Isolated by Ainsworth *et al.*,¹ the antibiotic was shown to be chemotherapeutic and selectively active against Gram-negative pathogens. It attracted attention in the course of screening operations by reason of its high activity against *Haemophilus pertussis* and *Eberthella typhosa*. Subsequently, its chemotherapeutic and pharmacological properties were described by Brownlee and Bushby.² They described its bactericidal nature and emphasized the difficulty with which resistant strains emerged. After the paper of Brownlee and Bushby had been lodged for publication, the report on "Polymyxin" by Stansly *et al.*³ became available. "Polymyxin," as described, is produced by another strain of the organism which produces "Aerosporin" and appears to have an identical antibacterial spectrum. However, as is apparent from the report, it differs both in antibacterial efficiency and in certain pharmacological properties from polymyxin A. There is an even earlier report of the antibacterial nature of crude liquid cultures of *B. polymyxa* presented by Benedict and Langlykke⁴ which anticipated by three days a description by Ainsworth *et al.*¹ of the chemotherapeutic activity of relatively pure material.

The simultaneous development of allied antibiotics in America and Britain produced by a common bacterial species is no accident. The process was catalyzed by the discovery of penicillin and by the fundamental work which uncovered streptothricin and streptomycin, and, in turn, it reflects the intense activity in this productive field. The preliminary clinical report of Swift⁵ indicated that the new antibiotic appeared to be effective in aborting early cases of pertussis, and in modifying the course of the disease in later cases in which mixed infections were already present. Simultaneously with the development of the pharmacology and chemotherapy of the material described as "Aerosporin," the studies on its isolation and purification (by Catch and Friedmann⁶) were published, as well as those on its chemical nature by Catch and Tudor Jones⁷ and by Tudor Jones.^{8, 9} These authors demonstrated the principally polypeptide nature of the antibiotic and showed that there was in its make-up leucine, threonine, and α, γ -diaminobutyric acid.

It was early observed that even "pure" samples of the antibiotic appeared

to be contaminated by a nephrotoxic principle (Brownlee and Bushby²). This problem was studied by Brownlee and Short,¹⁰ and it was shown that the lesions were restricted to the distal portion of the renal convoluted tubules and that their extent could be estimated histologically and, also, by observing the resultant proteinuria in rats or dogs. The damage was found to be similar to that due to *D*-serine (Artom, Fishman, and Morehead¹¹). It was found, further, that the substances which antagonized the nephrotoxic action of *D*-serine also antagonized the nephrotoxic activity of the principle in the new antibiotic. While it had been demonstrated clinically that "Aerosporin" could be successfully used in spite of its renal-damaging factor, this difficulty was an obvious defect. In clinical practice, methionine was observed to modify the proteinuria but not to eliminate it. Efforts to eliminate the toxic factor took the form of purification, identification of the site of damage and the development of counter measures, and strain selection to eliminate its formation. The third of these procedures proved to be the most productive and quickly provided a second antibiotic, which differed chemically from both "Polymyxin" and "Aerosporin" and proved to be free from nephrotoxic factor when examined both experimentally and clinically. Evidence will be presented later at this conference to justify these comments.

There was a second and not altogether unexpected outcome of strain selection. Knowledge was quickly built up of a number of antibiotics of essentially similar biological and chemical structure, pharmacological differences being noted from one to the other. Recognition of these diverse antibiotics, evidence for the existence of which will also be presented at this conference, raises in an acute form the question of nomenclature. A number of publications have appeared describing the isolation, chemotherapy, pharmacology, bacteriology, and clinical study of the antibiotic described as "Polymyxin." Further evidence to be presented here will demonstrate that this antibiotic is a specific entity differing from its allied substance previously described as "Aerosporin." The latter, for its part, also has been identified in the literature chemotherapeutically, pharmacologically, chemically, and clinically, as a specific entity differing from "Polymyxin." We now propose, at this conference, to present evidence for the existence of at least two more antibiotics of this type.

TABLE 1 compares the qualitative differences in the amino-acids derived by hydrolysis from the polypeptide component of four different antibiotics derived from *B. polymyxa*.

Whether the antibiotic studied at Peoria,⁴ that studied by Reese and Eisenberg,¹² and that studied by others, is "Polymyxin," "Aerosporin," or one of the other members, remains to be seen.

These are some of the considerations which make the acceptance of a generally approved generic name for this group of antibiotics a matter of immediate urgency. The name proposed is Polymyxin.*

* The name actually proposed at the conference was Bacillosporin. Subsequent discussion proved it to be unacceptable to the American workers who preferred the taxonomically derived Polymyxin. This was accepted by the British group in the interests of early agreement in a rapidly expanding field of research. It was agreed that Bacillosporin A (the original 'Aerosporin'), B, and C should be renamed Polymyxin A, B, and C and the original "Polymyxin" (Stansly *et al.*) be labeled D.

To adopt the excellent suggestion of Waksman,¹⁸ when discussing streptomycin, and to precede the generic name by modifying adjectives indicating the special points of difference would in this instance, and certainly at this time, produce names too unwieldy for practical use. The course which I suggest is the use of the alphabet. The equivalent one using cardinal numbers as suffixes is equally trivial and has a disadvantage of not being mnemonic. In default of an agreed nomenclature, the problem would become complex should further strains be isolated, yielding products of which the polypeptide portions of the molecule show further qualitative differences. In the same way, the determination of the quantitative compositions of the antibiotics would render the problem intractable.

TABLE 1

Poly-myxin	Culture No.	Leucine	Phenyl-alanine	Threonine	Serine	α,γ -Diamino-butyric acid
A	1984 2002 121	+	-	+	-	+
B	1419	+	+	+	-	+
C	2136 114J	-	+	+	-	+
D*	B71 (Lederle)	+	-	+	+	+

* Stansly *et al.*

Acknowledgments

A comprehensive series of studies of this kind draws upon the research of colleagues whose names do not appear on the title page. The following members of The Wellcome Research Laboratories have contributed to aspects of the research on the polymyxin group of antibiotics:

Antibiotics development and production, Mr. J. W. Woodthorpe; *isolation of new strains and study of fermentation conditions*, Mr. J. F. Spilbury; *bacterial and serological characters of strains*, Dr. A. E. Francis, Misses S. N. King and J. E. Rippon; *production and isolation*, Mr. D. M. Beattie, Misses B. M. Blanchard and M. P. Johnston, Messrs. P. S. S. F. Marsden, E. A. Terry, and J. W. Woodthorpe; *assay*, Miss M. F. Stevens and Mr. P. A. Young; *antibacterial and chemotherapeutic studies*, Misses J. M. Jacobs, M. M. Monk and S. M. Stewart; *pharmacological studies*, Mr. W. J. Miriam, Miss P. N. Wostear, and Mr. M. Woodbine; *purification and chemical studies*, Drs. W. M. Duffin, C. G. Pope, and C. G. Raison; and *electrophoretic studies*, Mr. H. V. Hempleman.

Literature Cited

1. AINSWORTH, G. C., A. M. BROWN, & G. BROWNLEE. 1947. 'Aerosporin,' an antibiotic produced by *Bacillus aerosporus* Greer. *Nature*. 160: 263.

2. BROWNLEE, G. & S. R. M. BUSHBY. 1947. Chemotherapy and pharmacology of aerosporin. A selective Gram-negative antibiotic. *Lancet* **254**: 127.
3. STANLEY, P. G., R. G. SHEPHERD, & H. J. WHITE. 1947. Polymyxin: a new chemotherapeutic agent. *Bull. Johns. Hopkins Hosp.* **81**: 43.
4. BENEDICT, R. G. & A. F. LANGLYKKE. 1947. Antibiotic activity of *Bacillus polymyxa*. *J. Bact.* **54**: 24.
5. SWIFT, P. N. 1948. Treatment of pertussis with aerosporin. *Lancet* **254**: 133.
6. CATCH, J. R. & R. FRILDMANN. 1948. The isolation and partial purification of 'Aerosporin.' *Biochem. J.* **42**: lii.
7. CATCH, J. R. & T. S. G. JONES. 1948. The chemical nature of 'aerosporin.' Part II. Isolation of a new natural amino-acid. *Biochem. J.* **42**: lii.
8. JONES, T. S. G. 1948. The chemical nature of aerosporin. *Biochem. J.*, **42**: xxxv.
9. JONES, T. S. G. 1948. The chemical nature of 'aerosporin.' III. The optical configuration of the leucine and threonine components. *Biochem. J.* **42**: lix.
10. BROWNLEE, G. & E. I. SHORT. 1948. Antagonism by amino-acids of renal tubule damaging substances present in 'Aerosporin' preparations. *Biochem. J.* **42**: liii.
11. ARTOM, G., W. H. FISHMAN, & R. P. MOREHEAD. 1945. The relative toxicity of *L* and *DL*-serine in rats. *Proc. Soc. Exp. Biol. & Med.* **60**: 284.
12. REISS, E. T. & G. M. EISENBERG. 1948. Polymyxin assay procedures. *Ann. N. Y. Acad. Sci.* **51** (5):
13. WAKSMAN, S. A. 1948. Nomenclature of streptomycin preparations. *Science* **107**: 233.

COMPARATIVE BIOLOGICAL STUDIES OF POLYMYXIN AND "AEROSPORIN"

By H. J. WHITE, C. M. ALVERSON, M. J. BAKER, AND
E. R. JACKSON

*Chemotherapy Division, Stamford Research Laboratories, American Cyanamid Company,
Stamford, Connecticut*

Polymyxin is a generic term which has been used¹⁻⁹ to designate non-homogeneous antibiotic material produced by *Bacillus polymyxa*. "Aerosporin" is the trademarked name introduced by Ainsworth, Brown, and Brownlee¹⁰ to designate a similar product from the same bacterial species.

A comparison of the original publications^{1,10} on polymyxin and "Aerosporin" strongly suggested that they were either identical or very closely related. Consequently, it became a matter of considerable interest to make a comparative study of the two substances to evaluate, by means of side-by-side tests, their points of similarity and difference. Such a study has been made possible through the mutual exchange of representative samples. The sample of "Aerosporin" (61 P 48) used in the present investigation was obtained from Dr. George Brownlee of the Wellcome Physiological Research Laboratories. It had an ascribed potency of 7940 "Aerosporin" units per milligram and, thus, was considered to be about 79 per cent "pure".¹¹ All comparative data have been based on the use of a sample of polymyxin (B71-91-2) with a potency of 1540 Stamford units per milligram, or a "purity" of approximately 77 per cent.¹

Both samples are mixtures of basic polypeptides and, although chemical studies are incomplete, a definite difference in their composition has already been firmly established.^{12,13} Hydrolysates of either polymyxin or "Aerosporin" contain leucine, threonine, α,γ -diaminobutyric acid, and an as yet unidentified C₆ acid. In addition to these components, polymyxin, but not "Aerosporin," contains the amino acid serine.

In this investigation, polymyxin and "Aerosporin" (as represented by the samples described above) have been compared with reference to their *in vitro* activity under various test conditions, therapeutic effectiveness in experimental infections, and acute toxicity for mice. Since, for all practical purposes, the samples were of equivalent purity, all comparisons have been made on a weight-for-weight basis.

Methods

Agar Dilution Spot Plate Method. A series of plates were set up, for each drug, with the drug incorporated in 2 per cent blood infusion agar in concentrations covering the desired test range, on a two-fold scale. The plates were allowed to dry, with raised covers, for an hour at 37°C. and were then inoculated by spotting one drop of a 10⁻¹ dilution of a 20-hour broth culture on a marked section of each plate. In this way, a single series of plates was used for titrating each drug against as many as twenty different test

organisms, simply by spotting twenty cultures on appropriately marked sections of each plate. After 24 hours' incubation at 37°C., the activity of the drug, against each organism, was measured in terms of the smallest concentration per ml. of agar which completely inhibited growth.

Broth Dilution Method. Serial two-fold dilutions of a glass-filtered aqueous solution of each drug were made aseptically in sterile Trypticase soy broth in final volumes of 5 ml. Each tube of each series was then inoculated with 0.2 ml., containing the desired inoculum, obtained by appropriate dilution of a 20-hour broth culture of test organisms. After incubation for 24 hours at 37°C., inhibitory activity was expressed as the smallest concentration of drug per ml. of broth which completely prevented visible growth. One-ml. volumes from tubes in which growth was absent were then subcultured into 9 ml. of sterile broth to determine bactericidal endpoints.

Experimental Infections. Experimental infections produced by *Pasteurella* strain #310 and *Klebsiella* strain AD are standard infections for chemotherapeutic studies in this Laboratory. These strains are highly virulent for mice and their virulence is maintained by continual mouse passage, on a bi-weekly schedule. The infections are produced by intraperitoneal inoculation of appropriate dilutions of 4-6 hour blood broth cultures. Virulence titrations indicate that a single organism, or clump, constitutes a lethal dose, as determined by plate count. The untreated infections are invariably characterized by rapidly progressing peritonitis and bacteremia, terminating in death within 16-36 hours. To establish that death was caused by the original infection, cultures are made of heart blood from mice dying during therapeutic experiments. Surviving mice were held for a total of 21 days after infection. Deaths among animals treated with single doses of polymyxin or "Aerosporin" rarely occurred after the tenth day following infection. On this basis, survival for 21 days was considered to be presumptive evidence of cure.

The experimental infection with *Haemophilus pertussis* was produced by intracerebral inoculation of 14-16 gram mice with 100,000 bacilli in 0.03 ml. volume. The culture (strain 18323) and standardized procedure for this infection were kindly supplied by Miss F. L. Clapp and Miss D. Novotny, of the Lederle Laboratories.

In all cases, our comparisons based on dose-effect data have been evaluated quantitatively by use of the Simplified Method of Wilcoxon and Litchfield.¹¹

Antibacterial Activity IN VITRO

Reproducibility of in vitro Test Results. Polymyxin and "Aerosporin" were titrated simultaneously against four bacterial strains by means of the agar dilution spot plate method. These titrations were replicated twenty times. The test, on any one day, consisted of titrating each drug against each of the four strains, or a total of eight titrations per day. Results of these replicate tests are shown in TABLE 1. It is evident that the activity

endpoints for either drug varied as much as four-fold, against each organism, from test to test. It is also evident that the relative activity of the two drugs varied from polymyxin being twice as active, to one-half as active, as "Aerosporin." Under these conditions, one can only conclude that the two drugs have the same order of activity.

Effect of Inoculum Size. The effect of inoculum size on the relative bactericidal activity of polymyxin and "Aerosporin" against each of four

TABLE 1
VARIATION IN RELATIVE ACTIVITY OF POLYMYXIN AND "AEROSPORIN" FROM TEST TO TEST
in vitro

Test No.	<i>E. coli</i>			<i>E. typhosa</i>			<i>K. pneumoniae</i>			<i>Past. multocida</i>		
	M.E.C. (μ g./ml)		Activity Ratio:	M.E.C (μ g./ml)		Activity Ratio:	M.E.C. (μ g./ml.)		Activity Ratio:	M.E.C. (μ g./ml)		Activity Ratio
	PM	A		PM	A		PM	A		PM	A	
1	2	2	1	1	1	1	1	1	1	1	2	2
2	1	2	2	1	1	1	1	2	2	2	2	1
3	1	2	2	0 25	0 5	2	0 5	1	2	1	1	1
4	4	4	1	0 5	0 5	1	1	1	1	2	2	1
5	2	4	2	0 25	0 5	2	0 25	0 5	2	1	2	2
6	4	4	1	1	2	2	1	2	2	2	2	1
7	4	2	0 5	1	1	1	1	2	2	4	2	0 5
8	4	2	0 5	1	1	1	1	1	1	4	2	0 5
9	2	4	2	1	1	1	1	1	1	1	2	2
10	1	1	1	1	0 5	0 5	1	1	1	1	1	1
11	2	2	1	1	1	1	1	1	1	1	2	2
12	1	2	2	1	0 5	0 5	1	0 5	0 5	1	2	2
13	1	2	2	1	1	1	1	0 5	0 5	1	2	2
14	1	2	2	0 5	0 5	1	0 5	1	2	1	1	1
15	1	2	2	1	1	1	0 5	1	2	1	1	1
16	2	2	1	0 5	1	2	1	2	2	1	1	1
17	2	2	1	0 5	1	2	0 5	1	2	1	1	1
18	1	2	2	0 25	0 5	2	1	1	1	1	1	1
19	1	2	2	0 5	0 5	1	0 5	0 5	1	0 5	0 5	1
20	1	2	2	0 5	0 5	1	0 5	1	2	0 5	1	2

M.E.C. - Minimum effective concentration, or the smallest concentration of drug which completely inhibited growth as determined by the agar dilution spot plate method
PM Polymyxin, A - "Aerosporin"

strains was determined by means of the broth dilution method. The test organisms included strains of *E. coli*, *E. typhosa*, *Klebsiella pneumoniae*, and *Pasteurella multocida*. Typical results are illustrated by FIGURE 1. (On the basis of these tests, it was concluded that polymyxin and "Aerosporin" are similarly affected by variation of inoculum size.

Spectrum Tests. Polymyxin and "Aerosporin" were titrated in side-by-side tests against each of twenty bacterial species by means of the agar dilution spot plate method. Results of duplicate comparisons of this kind are shown in TABLE 2. When one considers the variation from test to test with a single drug (TABLE 1), it is evident that a four-fold difference between

any two drugs, in a single comparison, is not striking enough to be considered significant. Consequently, the results shown in TABLE 2 must be taken as

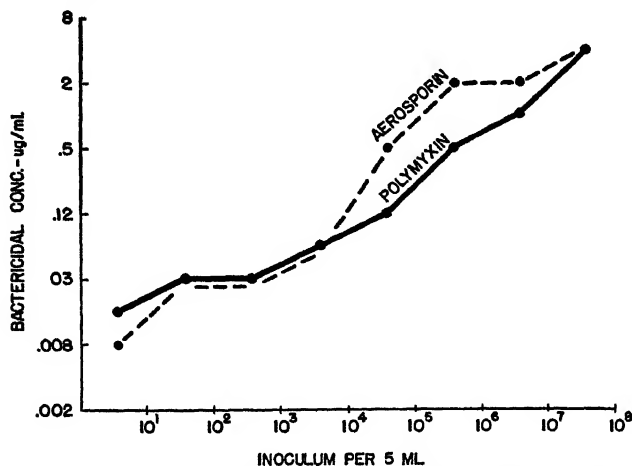


FIGURE 1 Inoculum size of *E. coli* (M) and bactericidal endpoint of drugs

TABLE 2
RELATIVE ACTIVITY OF POLYMYXIN AND "AEROSPORIN" *in vitro* (DUPLICATE TESTS)

Organism	Minimum Effective Concentrations $\mu\text{g./ml.}$		Activity ratio Polymyxin/ "Aerosporin"
	Polymyxin (as 77% pure)	"Aerosporin" (as 79% pure)	
<i>A. aerogenes</i> (S-1)	2, 2	8, 2	4, 1
<i>B. abortus</i> (19)	16, 16	32, 16	2, 1
<i>E. typhosa</i> (H)	0.5, 1	0.5, 1	1, 1
<i>E. coli</i> (M)	1, 2	1, 2	1, 1
<i>K. pneumoniae</i> (AD)	1, 1	1, 2	1, 2
<i>P. multocida</i> (310)	1, 1	1, 1	1, 1
<i>Ps. aeruginosa</i> (CM)	4, 4	2, 4	0.5, 1
<i>Sal. enteritidis</i> (149)	4, 8	1, 4	0.25, 0.5
<i>Sal. paratyphi</i> (S-1)	0.5, 0.5	0.5, 0.5	1, 1
<i>Sal. pullorum</i> (II)	1, 1	4, 4	4, 4
<i>Sal. schottmuelleri</i> (M)	2, 1	2, 2	1, 2
<i>Sh. flexner</i> (5733)	0.5, 1	1, 1	2, 1
<i>Sh. gallinarum</i> (469)	1, 1	1, 1	1, 1
<i>Sh. schmitz</i> (161)	0.5, 0.5	1, 2	2, 4
<i>Sh. shiga</i> (111)	1, 1	1, 1	1, 1
<i>Sh. sonnei</i> (151)	0.5, 0.5	1, 1	2, 2
<i>Pneumococcus</i> SVI	> 64, > 64	> 64, > 64	--
<i>Staphylococcus</i> 344	> 64, > 64	> 64, > 64	--
<i>Streptococcus</i> C203	> 64, > 64	> 64, > 64	--
<i>Streptococcus</i> B	> 64, > 64	> 64, > 64	--

Method: Agar dilution spot plate.

Medium: 2% blood infusion agar containing serial two fold dilutions of drugs.

Inoculum: One drop of 10⁻¹ dilution of 20 hour broth cultures; 20 species tested on each plate.

Incubation: 24 hours at 37°C.

evidence in support of the view that the two antibiotics have the same antibacterial spectrum.

Relative Activity in Agar Diffusion Tests. As shown in TABLES 1 and 2 and in FIGURE 1, when tested by the broth dilution or by the agar dilution method, polymyxin and "Aerosporin" were equally active against the M strain of *E. coli*. When the antibiotics were compared against the same strain by means of an agar diffusion method,³ an apparent difference in their activity was noted. With this method, the comparison was made by placing filter paper discs, saturated with various concentrations of each antibiotic, on agar plates seeded with *E. coli*. On a weight-for-weight basis, about three times as much "Aerosporin" as polymyxin was consistently needed to produce a given size of inhibition zone. Since the error of this particular method is approximately ± 15 per cent, the observed difference in behavior of the two substances was considered to be significant.

TABLE 3
BACTERICIDAL ACTIVITY OF POLYMYXIN AND "AEROSPORIN"

Drug in test mixture	Conc. $\mu\text{g}/\text{ml}$ in subculture plate	Colony counts per 0.2 ml. of test mixture					
		Polymyxin			"Aerosporin"		
		Initial	10 min	20 min	Initial	10 min	20 min.
6.4	.04	15	0	0	49	0	0
3.2	.02	47	0	0	84	3	0
1.6	.01	79	1	0	69	10	0
0.8	.005	83	7	5	85	49	12
0.4	.0025	71	36	21	66	68	59
0.2	.0013	76	56	48	53	76	75
0.1	.0006	79	83	76	77	80	87
0.05	.0003	78	85	71	69	70	68
0	0	64	70	85	70	65	90

Organism: *Klebsiella pneumoniae* strain AD

Medium: Trypticase soy broth; serial two-fold dilutions of each drug in 1 ml. volumes of broth.

Inoculum: 1 ml. of 10^{-8} dilution of 20-hour broth culture.

Temperature: 37°C.

Subculture: 0.2 ml. samples of each mixture plated in infusion agar at intervals of time indicated below. Colonies counted after 21 hours incubation

This apparent difference in the activity of polymyxin and "Aerosporin" may be due to the effect of different impurities on the diffusibility of active materials or, more probably, it may reflect a real difference in diffusibility of the active materials *per se*.

Bactericidal Action. The antibacterial action of polymyxin and "Aerosporin" is primarily bactericidal. This is readily demonstrated by the fact that viable cells cannot be subcultured from tubes which show no growth in the ordinary broth dilution test with these drugs. In addition, when large inocula (10^9 bacilli, or more) were plated in agar containing either drug, the absence of viable cells was repeatedly demonstrated by the failure to obtain growth when portions of the seeded agar were transplanted to broth. In these tests, the time of exposure of bacilli to drug was usually about 24 hours.

Results of an experiment which illustrates the rapidity of the lethal action of polymyxin or "Aerosporin" on the AD strain of *Klebsiella pneu-*

moniae are given in TABLE 3. In these experiments, relatively small numbers of bacilli were exposed to the drugs in order to make it possible to carry out plate counts at short intervals. It is evident from these results that test mixtures containing about 60 to 70 bacilli were sterilized within 20 minutes by 1.6 $\mu\text{g.}$ of either polymyxin or "Aerosporin" per ml. This type of experiment was repeated several times with similar results.

Bacteriolytic Action. In the course of tests to determine the relative speed of bactericidal action of polymyxin and streptomycin, it was noted that turbid suspensions of bacteria in polymyxin partially cleared, while those in streptomycin did not. This was further investigated by measuring the turbidity of suspensions of various organisms exposed to polymyxin, "Aerosporin," or streptomycin, for various periods of time and under various

TABLE 4
BACTERIOLYTIC ACTION OF POLYMYXIN AND "AEROSPORIN" AT 41°C

Time (minutes)	Turbidity readings			
	Polymyxin 64 $\mu\text{g.}/\text{ml.}$	"Aerosporin" 64 $\mu\text{g.}/\text{ml.}$	Streptomycin 2500 $\mu\text{g.}/\text{ml.}$	Control (buffer)
0	33	33	33	32
15	24	34	33	33
30	19	30	34	34
45	18	28	34	34
60	17	26	33	33
90	16	24	33	33
120	17	22	33	35
150	16	21	32	33
180	15	20	32	33
Decrease in Turbidity	55%	40%	0	0

Test Mixtures 5 ml of 128 $\mu\text{g.}/\text{ml.}$ (5000 $\mu\text{g.}$ streptomycin) of buffer solutions of antibiotics plus 5 ml of 6-hour Trypticase soy broth culture of *E. coli* (M)

Turbidity Readings Libbey photonreflectometer, instrument setting air at 23 $^{\circ}\text{F.}$, dist water at zero, glass standard at 37

conditions. Typical results are given in TABLE 4. It is apparent that, under exposure to polymyxin, the initial turbidity of the cell suspension rapidly decreased to a maximum of about 50 per cent. A similar decrease in turbidity was produced by "Aerosporin," while streptomycin failed to cause any change. This has been confirmed many times, with a variety of gram-negative strains, over a pH range of 6 to 8, a temperature range of 4 to 41°C., and in various suspending media, including agar. Heat-killed cells were not lysed by the drugs. Electron micrographs indicated that polymyxin-exposed cells become much less dense than normal cells.

Cross-Resistance. The relative sensitivity to polymyxin and "Aerosporin" of parent and polymyxin-resistant strains of *Pseudomonas aeruginosa* and *Klebsiella pneumoniae* was investigated by use of the broth dilution method. In these tests, replicate series of two-fold dilutions of each drug were made in 1-ml. volumes of Trypticase soy broth. The

inoculum for each tube of each series consisted of 0.2 ml. of a 10^{-2} broth dilution of a 22-hour broth culture of the appropriate strain. After 48 hours' incubation at 37°C., activity endpoints were taken as the smallest concentrations required for complete inhibition of growth.

Results of these tests are given in TABLE 5. It is evident that, within the limits of error of the method, the parent strains were equally sensitive to polymyxin and "Aerosporin." It is also evident that the polymyxin-resistant strains were similarly resistant to both antibiotics. These results indicate that polymyxin and "Aerosporin" are closely related, if not identical, substances with the same antibacterial mechanism of action.

Bacterial Resistance. In previous studies,¹ the relative incidence of variants resistant to either polymyxin or streptomycin was determined by seeding agar plates, containing drug, with large cell populations of various gram-negative strains. Under these conditions it was found that variants

TABLE 5
RELATIVE SENSITIVITY TO POLYMYXIN AND "AEROSPORIN" OF PARENT
AND POLYMYXIN-RESISTANT STRAINS

Strain	Polymyxin		"Aerosporin"	
	M E C μg./ml.	Sensitivity resistant strain parent strain	M E C μg./ml.	Sensitivity resistant strain parent strain
<i>Ps. aeruginosa</i> (parent)	4	—	1	—
<i>Ps. aeruginosa</i> (resistant)	2048	512	256	256
<i>K. pneumoniae</i> (parent)	0.5	—	0.5	—
<i>K. pneumoniae</i> (resistant #1)	256	512	512	1024
<i>K. pneumoniae</i> (resistant #2)	512	1024	1024	2048

M E C = Minimum effective concentration.

resistant to polymyxin could not be selected out of cell populations as large as 140 billion, while strains which were completely resistant to streptomycin could readily be obtained from relatively small populations of cells.

A preliminary comparison of this kind was made with polymyxin and "Aerosporin" by seeding seven billion typhoid bacilli (Hopkins strain), as determined by plate counts, into each of twenty agar pour plates; ten of these plates contained 10 μg. of polymyxin per ml. of agar and the remaining plates contained an equivalent concentration of "Aerosporin." Thus, 70×10^9 bacilli were exposed to a concentration of 10 μg./ml. in the case of each antibiotic. No resistant colonies appeared on any plate. This result confirms the previous finding that large populations of typhoid bacilli are free of variants resistant to polymyxin, as well as indicating that "Aerosporin" resembles polymyxin in this respect.

Experimental Infections

The relative effectiveness of polymyxin and "Aerosporin" was investigated in experimental infections produced by strains of *Pasteurella multocida*,

Klebsiella pneumoniae, and *Haemophilus pertussis* in mice. Experimental procedure is illustrated by TABLES 6 and 7, which give details of comparisons made with the *Pasteurella* infection.

Preliminary results with this infection (TABLE 6) indicated that polymyxin was slightly less active than "Aerosporin" on a dose-for-dose basis. Quantitative evaluation of the data, based on dose-effect curve parameters, however, showed that the difference observed could not be considered significant. In a more elaborate experiment (TABLE 7), the activity of

TABLE 6
A COMPARISON OF POLYMYXIN AND "AEROSPORIN" IN A *Pasteurella* INFECTION IN MICE
(EXPERIMENT # 1)

Dosage mg /kg	Survival on 21st day after infection					
	Polymyxin			"Aerosporin"		
	No alive No. tested	Per cent survival	Survival time ¹	No alive No. tested	Per cent survival	Survival time ¹
6.4	10/10	100		10/10	100	—
3.2	10/10	100	—	10/10	100	—
1.6	9/10	90	1.0	10/10	100	—
0.8	7/10	70	1.0	9/10	90	1.0
0.4	0/10	0	1.0	2/10	20	1.0
0.2	0/10	0	1.0	0/10	0	1.0

Dose-effect curve parameters†		
	Median survival dose	Slope function (S)
Polymyxin	0.60 (0.32-1.14) mg/kg	2.1 (1.1-4.2)
"Aerosporin"	0.53 (0.39-0.73) mg/kg	1.4 (1.1-1.8)

Activity ratio "Aerosporin" / Polymyxin = 1.1 (0.6-2.3)

Organism: *Pasteurella multocida* strain 310

Mice: Vanderwerken, 20 grams

Infection: Intraperitoneal, 0.5 ml of 10⁶ of 5 hour culture, 700 lethal doses

Treatment: Subcutaneous, one dose only, 0.2 ml volumes of drug solution, pH 6.1 administered immediately after infection

¹ Mean survival time, in days, for mice that died

† Values in parentheses indicate 19-20 confidence limits.

"Aerosporin" was found to lie between 1.1 and 1.6 times the activity of polymyxin, for odds of 19 to 1.

Results of similar quantitative comparisons of the two antibiotics in *Klebsiella* and *Haemophilus pertussis* infections are summarized in TABLE 8. It is evident that the activity of "Aerosporin" in the *Klebsiella* infection lies between equivalence of and 2.7 times the activity of polymyxin, for odds of 19 to 1.

With respect to the *H. pertussis* infection, it is apparent that "Aerosporin" was somewhat more effective on a dosage basis. In our hands, however, this particular infection has not been a satisfactory experimental

TABLE 7
A COMPARISON OF POMYMYXIN AND "AEROSPORIN" IN A *Pasteurella* INFECTION IN MICE
(EXPERIMENT #2)

Dosage mg /kg.	Survival on 21st day after infection					
	Polymyxin			"Aerosporin"		
	No alive No tested	Per cent survival	Survival time [†]	No. alive No. tested	Per cent survival	Survival time [*]
2.4	20/20	100	—	—	—	—
1.8	20/20	100	—	20/20	100	—
1.4	15/20	75	4.0	19/20	95	2.0
1.0	14/20	70	2.0	15/20	75	3.0
0.75	13/20	65	2.0	9/20	45	2.0
0.56	3/20	15	1.9	13/20	65	2.0
0.42	2/20	10	1.8	6/20	30	1.9
0.32	—	—	—	2/20	10	1.7

Dose-effect curve parameters [*]		
Median survival dose		Slope function (S)
Polymyxin		0.78 (0.68-0.90) mg./kg.
"Aerosporin"		0.59 (0.51-0.69) mg./kg.
		1.5 (1.3-1.7)
		1.7 (1.4-1.9)

$$\text{Activity ratio: } \frac{\text{"Aerosporin"}}{\text{Polymyxin}} = 1.3 (1.1-1.6)$$

Organism: *Pasteurella multocida*; strain 310.
Mice: Vandrwerken, 19-23 grams.
Infection: Intraperitoneal; 0.5 ml. of 10⁻⁶ of 5-hour culture; 600 lethal doses.
Treatment: Subcutaneous; one dose only; 0.2 ml. volumes of drug solutions at pH 6.4 administered immediately after infection.
^{*} Mean survival time, in days, for mice that died.
[†] Values in parentheses indicate 19/20 confidence limits.

TABLE 8
RELATIVE ACTIVITY OF POLYMYXIN AND "AEROSPORIN" IN EXPERIMENTAL INFECTIONS IN MICE (SUMMARY)

Infection	Test No.	Median effective dose mg./kg.		Relative activity	
		Polymyxin	"Aerosporin"	Poly- myxin	"Aerosporin"
<i>Pasteurella multocida</i>	1	.60(.32-1.1)	.53(.39-.73)	1.0	1.1(.6-2.3)
<i>Pasteurella multocida</i>	2	.78(.68-.90)	.59(.51-.69)	1.0	1.3(1.1-1.6)
<i>Klebsiella pneumoniae</i>	3	.52(.43-.62)	.35(.24-.50)	1.0	1.5(1.0-2.2)
<i>Klebsiella pneumoniae</i>	4	.66(.57-.76)	.28(.25-.32)	1.0	2.4(1.9-2.7)
<i>Klebsiella pneumoniae</i>	5	.50(.40-.63)	.26(.21-.33)	1.0	1.9(1.4-2.6)
<i>Haemophilus pertussis</i>	6	21 (12-37)	6 (4-10)	1.0	3.5(1.6-7.5)

Figures in parentheses indicate 19/20 confidence limits.

device for quantitative comparisons, since the dose-effect curves were rather flat and carried wide limits of error. This is reflected in the wide

TABLE 9
RELATIVE ACUTE TOXICITY OF POLYMYXIN AND "AEROSPORIN" FOR MICL

Route	Dose mg /kg.	Polymyxin		"Aerosporin"	
		No. dead No. tested	Per cent mortality	No. dead No. tested	Per cent mortality
Intraperitoneal	140	10/10	100	—	—
	100	19/20	95	—	—
	75	9/20	45	—	—
	56	0/10	0	10/10	100
	42	0/10	0	13/20	65
	32	—	—	3/20	15
	24	—	—	0/10	0
	18	—	—	0/10	0
L D. ₅₀ (19/20 confidence limits)		77 (70-85) mg /kg		39 (35-44) mg /kg	
Intravenous	24	10/10	100	—	—
	18	11/20	55	—	—
	14	2/20	10	10/10	100
	10	0/10	0	15/20	75
	7.5	0/10	0	4/20	20
	5.6	—	—	0/10	0
	4.2	—	—	0/10	0
L D. ₅₀ (19/20 confidence limits)		18 (16-20) mg /kg.		8.7 (8.0-9.5) mg./kg	
Subcutaneous	560	10/10	100	—	—
	420	19/20	95	—	—
	320	15/20	75	—	—
	240	12/20	60	10/10	100
	180	4/20	20	19/20	95
	140	1/10	10	20/20	100
	100	—	—	18/19	95
	75	—	—	12/19	63
	56	—	—	2/10	20
L D. ₅₀ (19/20 confidence limits)		230 (200-260) mg./kg.		68 (61-76) mg /kg.	

TABLE 10
RELATIVE ACUTE TOXICITY OF POLYMYXIN AND "AEROSPORIN" FOR MICL (SUMMARY)

Route	Median lethal dose mg./kg.		Relative toxicity	
	Polymyxin	"Aerosporin"	Polymyxin	"Aerosporin"
Intravenous	18 (16-20)	9 (8-10)	1.0	2.0 (1.9-2.3)
Intraperitoneal	77 (70-85)	39 (35-44)	1.0	2.0 (1.8-2.2)
Subcutaneous	230 (200-260)	68 (61-76)	1.0	3.4 (2.8-4.1)

Figures in parentheses indicate 19/20 confidence limits.

margin of error around the mean activity ratio for "Aerosporin" and polymyxin as shown in TABLE 8.

Relative Toxicity of Polymyxin and "Aerosporin"

The relative lethal toxicity of polymyxin and "Aerosporin" when the drugs were administered by the intravenous, intraperitoneal, or subcutaneous routes was determined by Vanderwerken Swiss mice. The groups of mice on each drug were equivalent with respect to weight and sex. Each animal was weighed individually and dosed on a body weight basis. The quantitative comparison for each route of administration was based on dosage-mortality curves, derived from graded doses of each drug, with groups of ten or twenty mice per dose. Results of these experiments are shown in detail in TABLE 9 and are summarized in TABLE 10.

Following intravenous or intraperitoneal administration, it is obvious that "Aerosporin" was twice as toxic as polymyxin on a dose-for-dose basis. In the case of intravenous dosage, deaths with either drug occurred within five minutes after injection. After intraperitoneal doses, deaths occurred up to three hours after injection.

When the drugs were given subcutaneously, "Aerosporin" was found to be from three to four times as lethal as polymyxin. Administration of either drug by this route was followed by deaths up to the sixth day. Delayed deaths occurred more frequently in the case of polymyxin and, in general, survival time was somewhat longer with this drug.

Summary and Conclusions

(1) Within the limits of error of the test methods employed, representative samples of polymyxin and "Aerosporin" were found to be essentially equivalent in their activity against a wide variety of gram-negative pathogens, *in vitro*.

(2) Both antibiotics were rapidly bactericidal and both were bacteriolytic in their action against certain sensitive strains.

(3) Polymyxin-resistant strains were equally resistant to "Aerosporin."

(4) On a weight-for-weight basis, polymyxin was about 30, 50, and 90 per cent as effective as "Aerosporin" in experimental infections produced in mice by strains of *Haemophilus pertussis*, *Klebsiella pneumoniae*, and *Pasteurella multocida*, respectively.

(5) Depending upon the route of administration in mice, "Aerosporin" was from two to three times as toxic as polymyxin.

Since no qualitative difference in the biological activity of representative samples of polymyxin and "Aerosporin" could be demonstrated, it is concluded that these substances are closely related members of the polymyxin-type group of antibiotics.

On the basis of all experimental data available, these closely related antibiotic substances appear to be equivalent with respect to their potentialities as useful chemotherapeutic agents.

Bibliography

1. STANSLY, P. G., R. G. SILLPHARD, & H. J. WHITE. 1947. Polymyxin: A new chemotherapeutic agent. Bull. Johns Hopkins Hosp. 81:43.
2. STANSLY, P. G. & M. E. SCHLOSSER. 1947. Studies on polymyxin: Isolation and

- identification of *Bacillus polymyxa* and differentiation of polymyxin from certain known antibiotics. J. Bact. 54: 549.
- 3 STANSLY, P. G. & M. E. SCHLOSSER. 1947. Studies on polymyxin: An agar diffusion method of assay. J. Bact. 54: 585.
 - 4 STANSLY, P. G. & N. H. ANANENKO. 1947. Resistance of polymyxin to some proteolytic enzymes. Arch. Biochem. 15: 473.
 - 5 BENEDICT, R. G. & F. H. STODOLA. 1948. Modification of an agar diffusion method of assay for polymyxin. J. Bact. 55: 286.
 - 6 STANSLY, P. G., M. E. SCHLOSSER, N. H. ANANENKO, & M. H. COOK. 1948. Studies on polymyxin: The production of fermentation liquor. J. Bact. 55: 573.
 - 7 SCHOENBACH, E. B., M. S. BRYER, E. A. BLISS, & P. H. LONG. 1948. Polymyxin. J.A.M.A. 136: 1096.
 - 8 BLISS, E. A. 1948. Studies on polymyxin. J. Bact. Abstracts.
 - 9 SHEPHERD, R. G., P. G. STANSLY, R. WINTERBOTTOM, J. P. ENGLISH, C. E. FELLOWS, N. H. ANANENKO, & G. L. GUILLET. 1948. Chemical studies on polymyxin. I. Isolation and preliminary purification. J. Am. Chem. Soc. 70: 3771.
 - 10 AINSWORTH, G. C., A. M. BROWN, & G. BROWNLEE. 1947. "Aerosporin," an antibiotic produced by *Bacillus aerosporus* (Greer). Nature, 160: 263.
 - 11 BROWNLEE, G. & S. R. M. BUSHBY. 1948. Chemotherapy and pharmacology of "Aerosporin." Lancet 254: 127.
 - 12 BELL, P. H., J. F. BONE, J. P. ENGLISH, C. E. FELLOWS, K. S. HOWARD, M. M. ROGERS, R. G. SHEPHERD, R. WINTERBOTTOM, A. C. DORNBUSH, S. KUSHNER, & Y. SUBBAROW. 1949. Chemical studies on polymyxin: comparison with "Aerosporin." Ann. N. Y. Acad. Sci. 51(5): 897-908.
 - 13 CATCH, J. R., T. S. G. JONES, & S. WILKINSON. 1949. Comparative chemical studies of polymyxin and "Aerosporin". Ann. N. Y. Acad. Sci. 51(5): 917-923.
 - 14 WILCOXON, F. & J. T. LITCHFIELD, JR. 1949. A simplified method of evaluating dose-effect experiments. J. Pharmacol. In press.

COMPARATIVE BIOLOGICAL STUDIES OF POLYMYXIN A AND POLYMYXIN D

By GEORGE BROWNLEE, S. R. M. BUSHBY, AND E. I. SHORT

The Wellcome Physiological Research Laboratories, Beckenham, Kent, England

Polymyxin A ('Aerosporin') is an antibiotic produced by a strain of *Bacillus polymyxa*, isolated from soil and air by Ainsworth, Brown, and Brownlee.¹ They noted its bactericidal action, its reluctance to produce resistant strains, and its chemotherapeutic activity in experimental animal infections due to *Haemophilus pertussis*, *Eberthella typhosa* and *Escherichia coli* of human origin, and *Brucella bronchiseptica* and *Escherichia coli* of veterinary origin. Subsequently, Brownlee and Bushby² reported a comprehensive chemotherapeutic and pharmacological study of polymyxin A, and Swift,³ in a preliminary note, referred to its successful clinical application to pertussis.

In the United States, Stansly, Shepherd, and White⁴ independently described the isolation of polymyxin D ('Polymyxin') from the fermentation liquors of *B. polymyxa* and reported its antibacterial, chemotherapeutic, and pharmacological properties. Also independently came the report of the discovery of the antibacterial activities of crude liquors in which *B. polymyxa* was grown by Benedict and Langlykke.⁵

In an accompanying paper (Brownlee⁶), there is reported an agreement between the workers involved on the revised nomenclature of antibiotics derived from *B. polymyxa*. The original "Aerosporin"¹⁻³ becomes polymyxin A, and two new distinct entities, described at this conference, become B and C; 'polymyxin' of Stansly *et al.*⁴ becomes polymyxin D.

A comparison of the reports on polymyxin A and polymyxin D show the antibiotics to be similar, yet demonstrate several apparent differences. We have, therefore, welcomed the opportunity, made possible by the receipt of 600 mg. of polymyxin from the Lederle Laboratories Division of the American Cyanamid Company, of making a comparative study of these two antibiotics.

The polymyxin D hydrochloride (strain B.71, 1350 units per mg.) was 67.5 per cent pure. Stansly *et al.*⁴ ascribed an arbitrary potency of 2,000 units per mg. to the pure salt. The polymyxin A hydrochloride specimens were of varying purity and are expressed in terms of "polymyxin A standard 1947," of 10,000 units per mg., a hydrochloride of 'near' chemical purity. The two units are, of course, unrelated.

Antibacterial Activity. The minimum concentration of the two antibiotics which inhibited the growth of a wide range of Gram-negative organisms was measured on solid medium using inocula from 24-hour broth cultures. Dilution in saline of the antibiotics was first made in geometrical two-fold dilutions, at 10 times the final concentration. These dilutions were then mixed in 1.5 ml. quantities with 13.5 ml. of nutrient agar con-

taining 0.5 per cent fresh horse blood, at 50°C., and then poured into 9 cm. Petri dishes. The plates when set were dried in an incubator at 37°C. for one hour and then inoculated with the test organisms. Inoculation with a straight wire was of single streaks of undiluted culture, thus enabling 17 organisms to be observed on the same plate. The spreading growths of *Ps. aeruginosa* and *P. vulgaris* were compared in tubes.

The activity of polymyxin D hydrochloride (strain B.71) was compared in this way with 5 different batches of polymyxin A hydrochloride. Two of them, P. 14 and TC 3, were early batches prepared some 12 months previously and both of 9 per cent purity. Batch 245 P 47 of 70 per cent purity

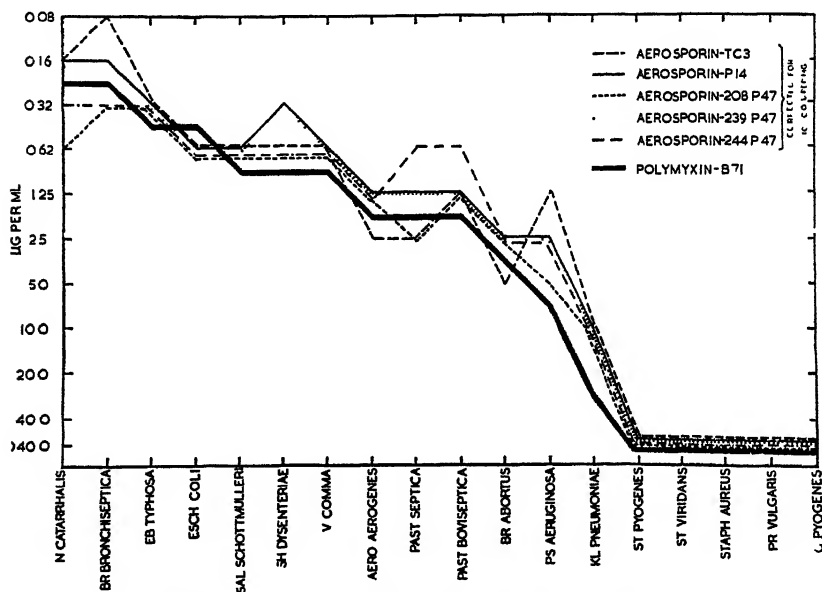


FIGURE 1. A comparison of the minimal concentrations in blood agar of one sample of polymyxin D (Strain B.71 Lederle) and five samples of polymyxin A ('Aerosporin') required to inhibit the growth of large inocula of 17 different bacterial organisms. The antibacterial activities of both antibiotics, shown in terms of pure substances, are similar.

and batch 208 P 47 of 100 per cent purity were recent, and 244 P 47 was a formaldehyde-bisulfite complex of 30 per cent purity calculated as hydrochloride.

The results (FIGURE 1) show polymyxin D and polymyxin A ('Aerosporin') to be similar, since the errors introduced by the test appear to be responsible for as much variation between the samples of polymyxin A as between any one of them and polymyxin D.

The close similarity between the antibiotics is not sustained when a comparison is made using small inocula. Brownlee and Bushby² showed the activity of polymyxin A against *E. typhosa* to depend upon the number of organisms exposed; polymyxin D differs in being less affected by the size of inoculum. The effect of varying the inocula of *E. typhosa* from 10^2

to 10^3 organisms in 5 ml. of nutrient broth with 4 samples of polymyxin A and polymyxin D (strain B. 71) is shown in FIGURE 2.

Both polymyxin A and polymyxin D may be assayed only by biological methods. The agar-cup penetration method has not proved satisfactory for accurate estimates of polymyxin A, for which purpose a serial dilution method using a sensitive organism is preferred. For the purpose of assay, the standard of reference is "polymyxin A Standard 1947" containing 10,000 units per mg. Attempts to define the activity of polymyxin D in terms of polymyxin A are defeated by the different behavior in sensitivity of the two substances to inoculum size. With small inocula of *E. typhosa*, polymyxin D has approximately 1000 polymyxin A units per mg., and with large inocula of *E. coli* 5000 polymyxin A units per mg.

In vivo Activity. Polymyxin A gives good protection to mice infected intracerebrally with mouse-passaged *H. pertussis*. This activity has been

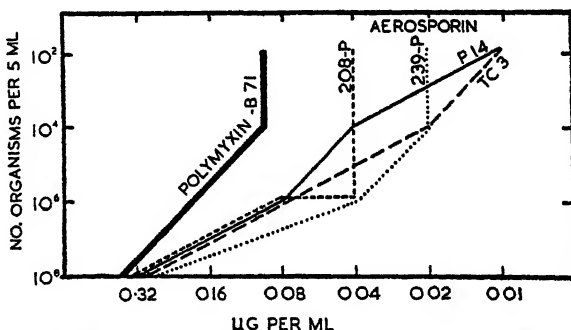


FIGURE 2. A comparison of the minimal concentrations in liquid media of one sample of polymyxin D (Strain B. 71 Lederle) and four samples of polymyxin A ('Aerosporin') required to inhibit varying numbers of *E. typhosa*. With large numbers of organisms the effects are similar, while with small numbers of organisms polymyxin A may be 2.5 to 10 times more active.

compared with that afforded by polymyxin D. Groups of mice (Schofield) infected with 0.05 ml. broth suspension containing 10^7 organisms from a 48-hour culture on Bordet-Gengou medium and consisting approximately of 10,000 average lethal doses, die within 4-7 days. Polymyxin A and polymyxin D were administered immediately after injection and again six hours later, and on the following two days by doses administered at 9:30 a.m. and 5:30 p.m. The results of these experiments covering a period of 15 days are shown in TABLE 1, where daily survivors are recorded.

Both drugs are seen to give protection, but, weight for weight, 10 times as much polymyxin D hydrochloride (strain B. 71) is required to protect as effectively as polymyxin A, or 6.75 times as much when corrected for "pure" polymyxin A of 2000 units per mg.

Both antibiotics give excellent protection to mice infected with *E. typhosa*. Mice infected intraperitoneally with 1,000 lethal doses of *E. typhosa* suspended in 0.5 ml. 5 per cent mucin were treated immediately, 6 hours later, and twice daily for the following 3 days with subcutaneous injections of

varying doses of the antibiotics. The results of these experiments covering a period of 7 days are shown in TABLE 2.

In this experiment, about twice as much polymyxin D (1350 u./mg.) appears to be required to afford the same degree of protection as pure polymyxin A.

TABLE 1
A COMPARISON OF POLYMYXIN A AND D IN MICE INFECTED WITH 10,000 LETHAL DOSES OF *H. pertussis*
(Antibiotics given parenterally twice daily for 3 days)

Drug	Dose mg.	No. of doses	No. of mice	No. mice surviving on day					Average survival rate, days
				3	6	9	12	15	
Polymyxin A 10,000 u./mg.*	0.05	6	15	15	14	14	13	13	13.8
	0.025	6	14	14	12	12	9	6	11.4
Polymyxin D 1350 u./mg.†	0.5	6	16	16	16	16	14	14	14.2
	0.2	6	14	14	13	9	7	5	10.3
	0.1	6	15	15	14	9	6	5	9.8
Lethal controls	—	—	14	14	8	—	—	—	4.7

* 10,000 units assumed "pure" (Brownlee and Bushby²).

† 1,350 units 67.5 per cent "pure" (Stansly *et al.*⁴).

TABLE 2
A COMPARISON OF POLYMYXIN A AND D IN MICE INFECTED WITH 1,000 LETHAL DOSES OF *E. typhosa*
(Antibiotics given parenterally twice daily for 3 days)

Drug	Dose mg.	No. of organisms	No. of mice	No. mice surviving on day				Average survival rate days
				1	3	5	7	
Polymyxin A 10,000 u./mg.*	0.1	5×10^7	18	17	17	17	17	6.6
	0.05	5×10^7	18	9	9	9	8	3.4
Polymyxin D 1,350 u./mg.†	0.3	5×10^7	12	12	12	11	11	6.9
	0.15	5×10^7	12	12	9	9	9	5.6
Lethal controls		5×10^7	18	1	0	—	—	0.0
		5×10^6	6	4	0	—	—	1.0
		5×10^4	6	3	3	3	3	3.5

* 10,000 units assumed "pure" (Brownlee and Bushby²).

† 1,350 units 67.5 per cent "pure" (Stansly *et al.*⁴).

Groups of mice, infected intraperitoneally with 10,000 and 1,000 lethal doses of *K. pneumoniae*, suspended in 0.5 ml. 5 per cent mucin, were treated immediately, 6 hours later, and twice daily for the following 3 days with subcutaneous injections of varying doses of the antibiotics. Details are shown in TABLE 3. Again, as in the case of the typhoid experiment, about twice the equivalent of pure polymyxin appears to be required to afford the same degree of protection as pure polymyxin A.

Toxicity. The intravenous acute LD₅₀ of polymyxin A, derived graphically from two groups of 10 mice, is 6.14 mg. per kg., and that of polymyxin D estimated simultaneously, 18.0 mg. per kg. The toxic signs of the two drugs also differ significantly. With lethal doses of polymyxin A, death occurs from respiratory failure in less than two minutes, and with near-lethal doses there are clonic convulsions followed by curare-like paralysis, associated with marked respiratory embarrassment and cyanosis, but the animals recover within ten minutes. With polymyxin D, recovery from near-lethal doses may take 30 minutes or longer, and the paralysis is absent. Death is also delayed, sometimes as much as 30 minutes.

All samples of polymyxin A, irrespective of purity, cause temporary

TABLE 3
A COMPARISON OF POLYMYXIN A AND D IN MICE INFECTED WITH 10,000 AND 1,000
LETHAL DOSES OF *K. pneumoniae*
(Antibiotics given parenterally for 3 days)

Drug	Dose mg.	No. of organisms	No. of mice	No. mice surviv- ing on day				Average survival rate days
				1	3	5	7	
Polymyxin A 10,000 u/mg.*	0.1	5 × 10 ⁸	6	5	1	1	1	1.8
		5 × 10 ⁷	6	5	2	2	1	3.8
	0.05	5 × 10 ⁸	6	6	3	3	3	4.0
		5 × 10 ⁷	6	6	3	3	3	4.0
Polymyxin D 1,350 u/mg.†	0.3	5 × 10 ⁸	6	6	2	1	1	2.5
		5 × 10 ⁷	6	6	2	2	2	3.3
	0.15	5 × 10 ⁸	6	6	1	1	1	2.5
		5 × 10 ⁷	6	6	3	3	2	2.5
Lethal controls		5 × 10 ⁸	6	0	—	—	—	0.0
		5 × 10 ⁷	6	0	—	—	—	0.0
		5 × 10 ⁶	6	2	2	2	2	3.0
		5 × 10 ⁴	6	4	4	4	4	5.0

* 10,000 units assumed "pure" (Brownlee and Bushby²).

† 1,350 units 67.5 per cent "pure" (Stansly *et al.*⁴).

oliguria in hydrated rats, when given in large doses. This antidiuretic effect is also produced by polymyxin D. The renal tubular damaging factor, which causes a transitory proteinuria, and which contaminates most samples of polymyxin A, was also present in polymyxin D.

Conclusions

A comparison of the biological properties of polymyxin D (Strain B. 71) and several preparations of polymyxin A of widely differing potencies show some similarities and several differences.

A side-by-side comparison of their antibacterial spectra shows a similarity within the limits of the method; polymyxin D appears to have about half the intrinsic potency of polymyxin A.

The similarity breaks down when a comparison of antibacterial activity

is made with small inocula. The activity of polymyxin A varies with inocula; polymyxin D is similarly, but less affected.

These differences defeat attempts to define polymyxin D in terms of polymyxin A. With small inocula, polymyxin D has about one-tenth the activity of polymyxin A; with large inocula, about one half.

Both antibiotics protect mice from many lethal doses of *H. pertussis*. Polymyxin A is about seven times as effective as polymyxin D.

Compared by their capacity to protect mice against many lethal doses of *E. typhosa*, or *K. pneumoniae*, polymyxin A is about twice as effective as polymyxin D.

In acute experiments, polymyxin D has one-third the acute toxicity of polymyxin A. Moreover, the toxic signs are characteristic for both drugs.

References

1. AINSWORTH, G. C., A. M. BROWN, & G. BROWNLEE. 1947. Aerosporin, an antibiotic produced by *Bacillus aerosporus* Greer. *Nature*, **160**: 263.
2. BROWNLEE, G. & S. R. M. BUSHBY. 1947. Chemotherapy and pharmacology of aerosporin. A selective Gram-negative antibiotic. *Lancet*, **254**: 127.
3. SWIFT, P. N. 1948. Treatment of pertussis with aerosporin. *Lancet* **254**: 133.
4. STANSLY, P. G., R. G. SHEPHERD, & H. J. WHITE. 1947. Polymyxin: a new chemotherapeutic agent. *Bull. Johns. Hopkins Hosp.* **81**: 43.
5. BENEDICT, R. G. & A. F. LANGLYKKE. 1947. Antibiotic activity of *Bacillus polymyxa*. *J. Bact.* **54**: 24.
6. BROWNLEE, G. 1948. Antibiotics derived from *Bacillus polymyxa*. *Ann. N. Y. Acad. Sci.* **51**(5): 875-878.

CHEMICAL STUDIES ON POLYMYXIN: COMPARISON WITH "AEROSPORIN"*

By P. H. BELL, J. F. BONE, J. P. ENGLISH, C. E. FELLOWS, K. S. HOWARD,
M. M. ROGERS, R. G. SHEPHERD, AND R. WINTERBOTTOM

*Chemotherapy Division, Stamford Research Laboratories, American Cyanamid Company,
Stamford, Connecticut*

and A. C. DORNBUSH, S. KUSHNER, AND Y. SUBBAROW

Lederle Laboratories Division, American Cyanamid Company, Pearl River, N. Y.

This paper is a preliminary report on the chemistry of polymyxin† and its degradation products. "Aerosporin"‡ has been compared with particular samples of polymyxin^{2, 3} at a number of points of this study and the results of this comparison are included.

The work is divided into two main phases. The first of these is the study of the homogeneity of these samples of polymyxin. The second is the examination of the chemistry of polymyxin and the determination of its constituent parts.

In studying the purity of this chemotherapeutic agent, three problems arise immediately. Is there more than one biologically active component? If so, can the mixture be separated? Having a material pure by bioassay, can chemically pure material be prepared? Affirmative answers can be returned to the first two questions, while the third must be left unanswered for lack of information.

As received from production,^{2, 3} polymyxin is in the form of a hydrochloride. It is a nearly colorless powder which assays, in Stamford units,⁴ about 1500 units per milligram and melts with decomposition at 228–230°C. (varying with rate of heating and degree of hydration). It is very soluble (more than 40 per cent) in water and methanol, and the solubility decreases in higher alcohols. It is levorotatory ($\alpha_D^{25} = -40^\circ$ ($C = 1.05$ in water)). It is insoluble in ethers, esters, ketones, hydrocarbons, and the chlorinated solvents. The birefringent base may be precipitated from concentrated aqueous solutions of the hydrochloride by saturation with ammonia. This form is slightly soluble in water and almost insoluble in alcohol, and decomposes at a higher temperature and over a range.

The material forms water-insoluble salts with a number of precipitants such as picric acid, helianthic acid, Reinecke salt, and the like. None of these was of a nature to permit their purification by conventional fractionation procedures such as crystallization or precipitation and, except for certain analytical purposes, they were without value for a study of the purity of the material.

It was found that the material was distributed between water and those

* "Aerosporin" is the trademark of Burroughs Wellcome Ltd. for an antibiotic from *B. polymyxa*.¹

† Polymyxin is the generic term for the antibiotics (first isolated from *B. polymyxa* culture filtrates) which have closely related biological² and chemical characteristics (described herein).

water-immiscible alcohols which dissolve water, and that the amount which enters the non-aqueous phase is in direct proportion to the amount of water dissolved therein. Other solvents did not extract the activity. The distribution coefficient is inversely proportional to the salt concentration of the aqueous phase and to the pH.

In using bioassay values for the calculation of distribution coefficients, exact values are not obtainable. The agar-diffusion assay method which has been used has an estimated error of ± 15 per cent. This means that a true distribution coefficient of 1 could have experimentally determined values of 1 ± 0.2 . TABLE 1 presents the distribution coefficient of polymyxin as determined in three solvent systems and values for "Aerosporin" in two of these.

It is clear that the polymyxin examined and "Aerosporin" differ significantly in this property. It is quite possible that such determinations would permit the ready differentiation of the two materials when carried out di-

TABLE 1
DISTRIBUTION COEFFICIENTS
(Concentration aq. phase/Concentration non-aq. phase)

n-BuOH +	Polymyxin	"Aerosporin"
pH 5 SS buffer*	13 -14	83
pH 5 SS buffer* + 5% NaCl	1.5- 1.8	6-7
pH 2 SS buffer*	40	

* 0.015 M Sulfosuccinate.

rectly on fermentation liquor. Such studies also lay the groundwork for partition chromatography and countercurrent distribution studies.

Partition chromatography was first applied to the material. In these studies, Hyflo Supercel has been used as the support for the aqueous phase (0.65-0.75 cc. of buffer per gram of support) as it does not adsorb polymyxin under these conditions and can be obtained in large uniform batches more conveniently than can the silica gel of the original publications.⁵ A comparison of polymyxin and "Aerosporin," using equal weights of each on two columns prepared in the same way, is shown in FIGURE 1.

Two facts are revealed by these curves. The first is that polymyxin is made up of more than one active material. The second is that "Aerosporin" is not identical with the main body of the polymyxin. That it is not the same as the trailing material is shown by the fact that the isolated material from the tail of the polymyxin curve assays 1800-1900 units per mg. *versus* the calculated 600 Stamford units per mg.⁶ of pure "Aerosporin". The essential difference of polymyxin and "Aerosporin" is also borne out by paper chromatography of the two materials when the R_f values of the antibacterial activity were obtained. In obtaining these results, the strips were dried after development, placed briefly on long *E. coli*-seeded agar plates and removed. This last operation minimizes diffusion of the active

material due to wetting the paper in this operation, thereby giving smaller zones of inhibition.

A similar method has also been used in reducing the labor involved in assaying the large number of cuts collected in the course of the chromatographic work. Each cut is spotted on a strip on filter paper, using a melting point capillary to deliver the sample. Dilutions of a standard are also spotted on strips and all the strips are plated as above, without, however, removing the strips. Using this method, it is possible to make thirty assays in duplicate on a plate 115 x 280 mm.

FIGURE 2 shows the results of the partition chromatography of a larger amount of polymyxin in the same system and on a larger column. The inhomogeneity is quite clear, and the first question as to the presence of

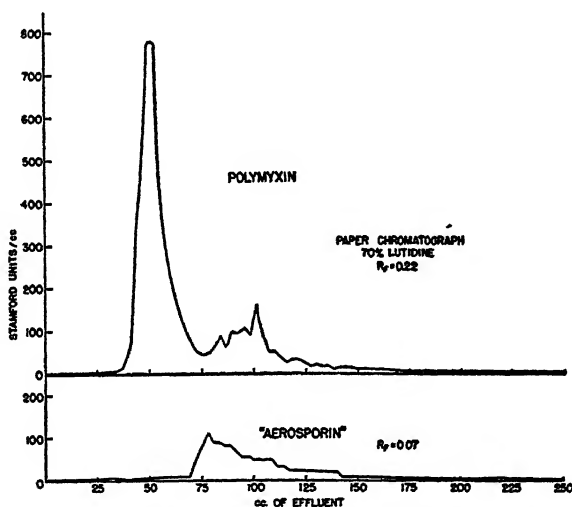


FIGURE 1. Partition Chromatograph. n -BuOH-pH₂ 0.015 M Sulfosuccinate Buffer on Hyflo.

more than one active component has been answered affirmatively. The question as to the bioassay purity of the material in the main peak was shown by quantitative isolation of the active substance and its rechromatography as presented in FIGURE 3. This shows that the more slowly moving activities of FIGURE 2 have been eliminated by the chromatographic treatment and, by comparison with columns on a similar scale, that the material behaves as does the main peak originally.

Further confirmation of the bioassay purity of the main peak material from FIGURE 2 is obtained by an examination of the material by counter-current distribution by the method of Craig.⁸ The results of such a distribution are shown in FIGURE 4, which shows the fraction of the total antibacterial activity in each tube. The calculated curve, within experimental error, fits the experimental curve. The great disparity in volumes of the two phases used was made necessary by the desire to have an effec-

tive distribution coefficient of about 1 without using large quantities of salt, which interfere with both the assay⁹ and isolation procedures.

The reality of the material occurring at 250–500 cc. in FIGURE 2 and its difference from the main peak is shown by its countercurrent distribution in the same system as that of FIGURE 4. This is illustrated in FIGURE 5,

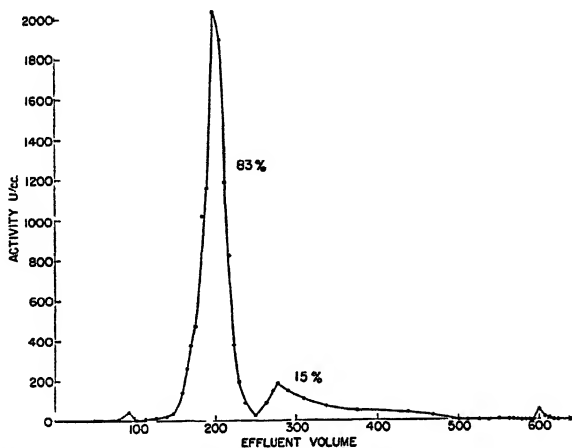


FIGURE 2. Partition Chromatography-Polymyxin. Hyflo Column (167). BuOH· pH₂ Buffer System, Recovery 91%.

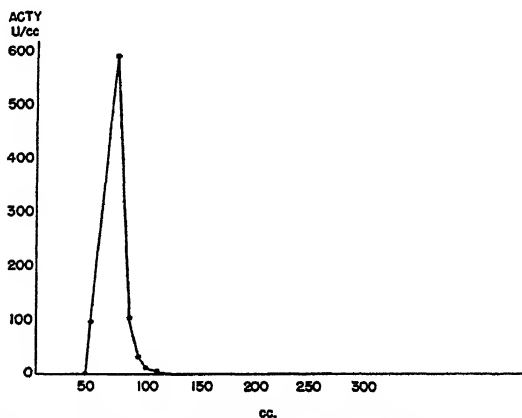


FIGURE 3. Repetition of Partition Chromatography on Main Peak—Column 167. BuOH· pH₂ Buffer. Hyflo Column. 92% Recovery.

revealing the different nature of the material. Rechromatography of this material confirms this conclusion. Unpublished data on countercurrent distribution of this sample of polymyxin also show the presence of such high-partition-coefficient activity.

At present, there is no information as to the chemical purity of the bioassay pure polymyxin fraction. Work is under way on this aspect of the problem.

At the same time that these studies of the number and purity of the components of polymyxin were under way, an investigation of the chemistry of 1500 unit per mg. material was begun. The pertinent findings of this investigation are summarized in TABLE 2.

The color reactions and the nature of the nitrogen in the substance early led to the conclusion that the material was a polypeptide. The agreement of total nitrogen values with the Van Slyke nitrogen after hydrolysis ex-

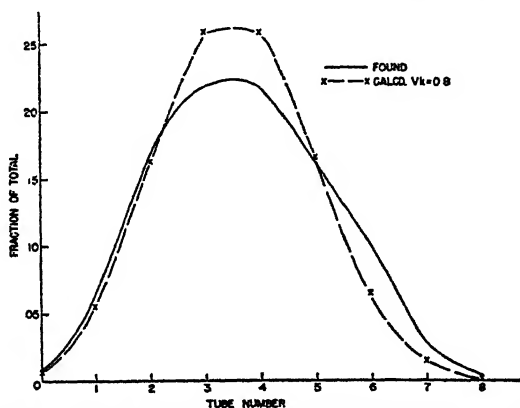


FIGURE 4. Main Peak. Craig-Type Distribution. $n\text{-BuOH:pH } 0.15 \text{ M Sulfosuccinate Buffer::16.5:1}$.

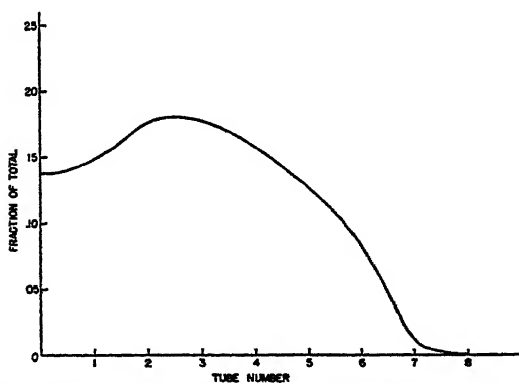


FIGURE 5. Tail Peak. Craig-Type Distribution. $n\text{-BuOH:pH } 5 \text{ Sulfosuccinate Buffer::16.5:1}$.

cludes the nitrogen-heterocyclic amino acids and arginine from consideration. The lack of a characteristic ultraviolet spectrum rules out the presence of the aromatic amino acids. The absence of sulfur does not permit the presence of the sulfur-containing amino acids. The water and formol titrations, taken in conjunction with the chloride values, and the lack of ninhydrin carboxyl show that the material is basic in nature and without acidic function. That the molecule consists of about four of these units is shown by the molecular weight values. The Signer¹⁰ value was determined in acetic acid using the free base and the osmometric value is by

a modified Baldes¹¹ method on the hydrochloride. The only assumption involved in the latter determination is that the hydrochloride is completely ionized. The helianthate value was calculated from the sulfur content of the salt, and the picrate value from the picric acid content as determined colorimetrically. The DNP (for 2,4-dinitrophenyl) value is the colorimetric equivalent weight of polymyxin which has been reacted with 2,4-dinitrofluorobenzene.¹² It is included here to show that there has been complete reaction with the basic groups.

TABLE 2
ANALYTICAL PROPERTIES OF POLYMYXIN

Base Equiv. Wt. on			Per Cent Nitrogen		
			Dumas	Van Slyke	Ninhydrin CO ₂
Cl ⁻	285	Intact	15.0	5 (3 min.)-7 (30 min.)	0
Picrate	286	Hydrolyzed		15.8 (3 and 30 min.)	ca. 10
Helianthate	277		No ammonia on hydrolysis		
Water titr.	300				
Formol titr.	306				
DNP	276				
		Molecular Wt. (Base)			
		1150 (Osmometer)		"Aerosporin" 1293 (Osmometer)	
		ca. 1000 (Signer)			

No characteristic UV spectrum, no sulfur, no alkoxy, no ash.
Positive ninhydrin and biuret, negative Sakaguchi.

TABLE 3
FATTY ACID CONSTITUENT

HCl Polymyxin → 12.6% acid (11.3% radical) (gravimetric)		
<i>Analysis</i>		
Acid	Calculated for C ₉ H ₁₈ O ₂ :	C, 68.4; H, 11.4
	Found	69.2 11.6
Amide (m.p. 92-93)	Calculated for C ₉ H ₁₉ NO:	C, 68.9; H, 12.1
	Found	68.9 12.3

With such knowledge of the molecule as a whole, attention was turned to the hydrolysis products. In addition to the nitrogenous products, there was present a nitrogen-free acidic substance. This was exhaustively extracted with petroleum ether and its properties are shown in TABLE 3. The data are in fair agreement for the proposed nine carbon acid. Comparison of the infrared absorption spectra of the acid, amide, and the methyl ester with pelargonic acid and derivatives¹³ shows that the material is not this compound. "Aerosporin" also has such an acidic material.

The residual hydrolysate was examined for the amino acids that the analytical data had indicated to be present. The relevant data are presented in TABLE 4. The presence of threonine and its orientation was recognized by microbiological assay. The α,γ -diaminobutyric acid was

identified both by isolation and as the γ -2,4-dinitrophenyl derivative obtained by hydrolysis of dinitrophenyl polymyxin. The hydantoin of the isolated material (m.p. 197°C.) did not depress the melting point (197°C.) of the synthetic.

The leucine and serine were identified by paper chromatography and the orientation of the former was determined on a sample isolated from paper chromatography on a large sheet.¹¹ The orientation of the serine was arrived at by the rotation of the isolated dinitrophenyl derivative. All

TABLE 4
IDENTIFICATION OF AMINO ACIDS

Amino Acid	Chromatographic Comparisons		Isolate
	Amino Acids (Paper)	DNP Derivatives (Partition)	
D-Leucine	7 systems	4 systems	Amino acid* (I.R., X-Ray)
L-Threonine†	7 systems	4 systems	Amino acid* (I.R., X-Ray)
D-Serine	7 systems	4 systems	DNP serine* (I.R.)
L- α , γ -Diamino-butyric acid	7 systems	4 systems	Amino acid* (I.R., X-Ray)
			α , γ -Di DNP (I.R.)
			γ -DNP HCl* (I.R., m.p.)
			Hydantoin (m.p.)

* Optical activity determined on this material.

† Microbiological assay.

TABLE 5
AMINO ACIDS IN POLYMYXIN BY PAPER CHROMATOGRAPHY

	Present in	Solvent systems
Leucine	Pm, As	1. AcOH·BuOH·H ₂ O (50:25:25) 2. Phenol
Threonine	Pm, As	3. AcOH·BuOH·H ₂ O (25:50:25) 4. Lutidine·BuOH·H ₂ O (50:25:25)
Serine	Pm	5. Lutidine·H ₂ O (70:30)
α , γ -Diaminobutyric acid	Pm, As	6. AcOH·BuOH·H ₂ O 7. BuOH·NH ₃

Pm = Polymyxin; As = Aerosporin.

of the isolated materials were also identified by comparison of infrared spectra¹³ and x-ray diffraction patterns.

The comparison of the amino acid compositions of polymyxin and "Aerosporin" is shown in TABLE 5. The findings listed here are all based upon side-by-side comparisons of the synthetic amino acids with the hydrolysates of polymyxin and "Aerosporin" in each of the solvent systems. The homogeneous material from the main peak of FIGURE 2 also shows the same acids. The solvent systems are listed in order of decreasing R_f of leucine in them. The validity of such an identification is strengthened by the observation that leucine and its six α -amino structural isomers are clearly

differentiated by this method. The following amino acids were excluded by direct comparison with the hydrolysate in one or another of these solvent systems: alanine, asparagine, aspartic acid, citrulline, glutamic acid, glycine, lysine, valine, and ornithine.

The identities of the amino acids were checked in two ways on isolates

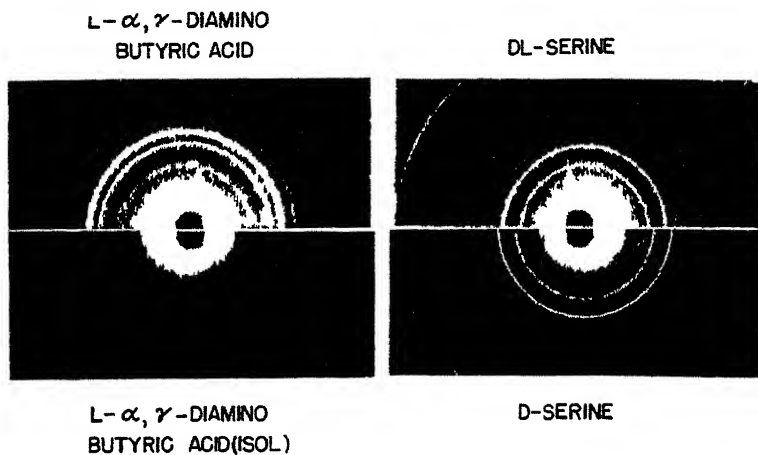


FIGURE 6.

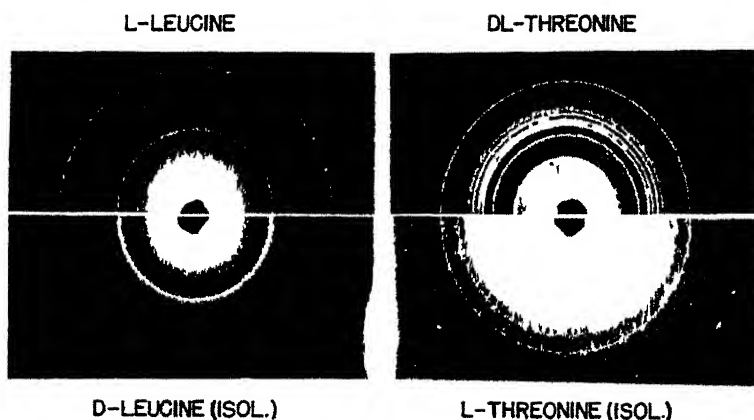


FIGURE 7.

made from one-dimensional paper chromatograms on large sheets. They were identified by comparison of their infrared spectra with those of known samples. Comparisons of the x-ray diffraction patterns of leucine, threonine, and α , γ -diaminobutyric acids with the corresponding isolated materials were made by a special technique.¹⁵ The results are shown in FIGURES

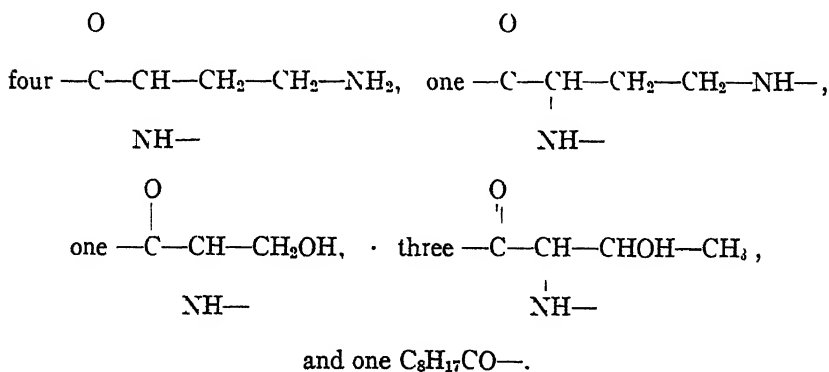
diaminobutyric acid (synthesized from *L*-glutamic acid^{1b}) both by comparison of rates on the partition column, by comparison of infrared spectrum¹³ melting point, and optical rotation. Unreacted α,γ -diaminobutyric acid was shown to be present in the hydrolysate by paper chromatography.

Since the dinitrophenyl compounds had high and uniform molecular extinction coefficients (ca. 17,000 when read at 3625 Å in 0.05M phosphate buffer at pH 7) they were used for quantitative studies on the ratios of the amino acids. This is shown in TABLE 6. The quantitative work has been carried out by hydrolysis of polymyxin followed by dinitrophenylation of the reaction mixture and separation of the components by partition chromatography. The amounts of each were then determined colorimetrically in the column effluents. The configurations of the serine and diaminobutyric acid components were shown by the rotation of the 2,4-dinitrophenyl derivatives. The amount and configuration of threonine in the hydrolysate was also shown by the response of a microbiological assay.

TABLE 7

	Calculated	Found
C	43.5	43.5
H	7.8	7.0
N	15.2	15.1
O	23.2	24.0 (difference)
Cl	10.3	10.5
Molecular weight (base)	1144	1150
Equivalent weight	286	275-290

With this information, it is possible to assign a tentative structure to the polymyxin molecule. There are available for this exercise the following numbers of the indicated residues:



Such an assemblage of residues indicates an empirical formula (for the hydrochloride) of $\text{C}_{50}\text{H}_{97}\text{N}_{15}\text{O}_{15}\text{Cl}_4$. The calculated values for the pentahy-

drate of this formula and those found are shown in TABLE 7. In view of the demonstrated inhomogeneity of the material, this agreement suggests that the various biologically active materials are very closely related in composition.

An arrangement of these residues which would be in agreement with the properties of polymyxin is sketched below.

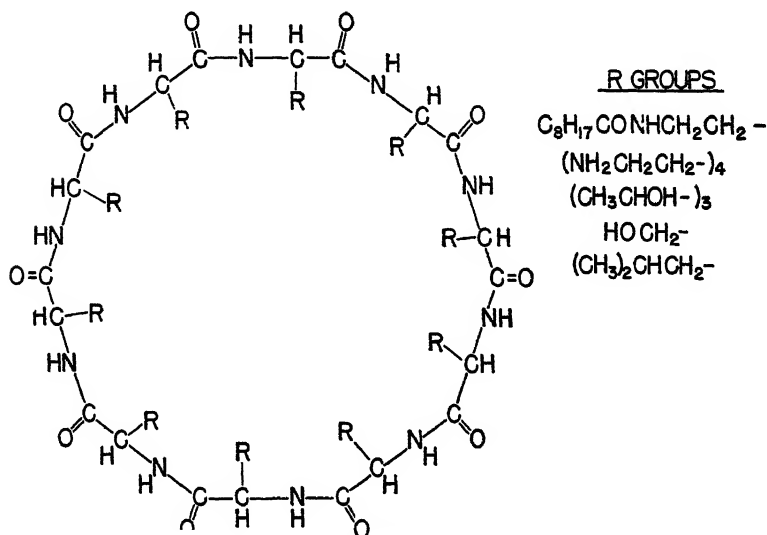


FIGURE 9.

This is similar to the closed cycle postulated for Gramicidin S.¹⁴

Natural hesitation about postulating any specific arrangement is understandable when it is realized that some 5,000 such possibilities exist.

In summary, it may be said that this sample of polymyxin is a basic polypeptide containing *L*- α , γ -diaminobutyric acid, *D*-serine, *L*-threonine, *D*-leucine, and a C_8 fatty acid. It differs from "Aerosporin" both in gross properties and in containing serine. Polymyxin has been postulated to be a cyclic material and evidence for a tentative molecular formula has been advanced.

Bibliography

1. AINSWORTH, G. C., A. M. BROWN, & G. BROWNLEE. 1947. *Nature* **160**: 263. The sample used in these studies was generously supplied by Dr. Brownlee and coded 61P48.
2. STANSLY, P. G., R. G. SHEPHERD, & H. J. WHITE. 1947. *Bulletin of the Johns Hopkins Hospital*. **81**: 43.
3. PORTER, J. N., H. KRUPKA, & R. BROSCARD. 1949. Production and isolation of polymyxin. *Ann. N. Y. Acad. Sci.* **51**(5): 857-865.
4. STANSLY, P. G. & M. E. SCHLOSSER. 1947. *J. Bact.* **54**: 585.
5. MARTIN, A. J. P. & R. L. M. SYNGE. 1941. *Biochem. J.* **36**: 1358.
6. WHITE, H. J., C. ALVERSON, M. J. BAKER, & E. R. JACKSON. 1949. *Ann. N. Y. Acad. Sci.*, **51**(5): 879-890.

7. CONSDEN, R., A. H. GORDON, & A. J. P. MARTIN. 1944. *Biochem. J.* **38**: 224.
8. CRAIG, L. C. 1944. *J. Biol. Chem.* **155**: 519.
9. SHEPHERD, R. G., P. G. STANSLY, R. WINTERBOTTOM, J. P. ENGLISH, C. E. FELLOWS, N. ANANENKO, & G. GUILLET. 1948. *J. Am. Chem. Soc.* **70**: 3771.
10. SIGNER, R. 1930. *Ann.* **478**: 246.
11. BALDES, E. J. 1939. *Biodynamica* **46**.
12. SANGER, F. 1945. *Biochem. J.* **39**: 507.
13. GORE, R. C. 1949. *Ann. N. Y. Acad. Sci.* **51**(5): 924-934.
14. CONSDEN, R., A. H. GORDON, A. J. P. MARTIN, & R. L. M. SYNGE. 1947. *Biochem. J.* **41**: 596.
15. CHRIST, C. L., M. BOTTY, & C. J. BURTON. 1948. *Science*. **108**: 91.
16. ADAMSON, D. W. 1939. *J. Chem. Soc.* 1564.

CHEMICAL EVIDENCE FOR THE MULTIPLICITY OF THE ANTIBIOTICS PRODUCED BY *BACILLUS POLYMYXA*

By TUDOR S. G. JONES

The Wellcome Chemical Research Laboratories, Beckenham, Kent, England

The announcement of the discovery of "Aerosporin" by Ainsworth, Brown, and Brownlee¹ and of "polymyxin" (Benedict and Langlykke,² Stanley *et al.*³), among the metabolic products of the organisms of the *B. aerosporus*-*B. polymyxa* group, raised the inevitable question of the chemical identity of these materials. The substance polymyxin A, formerly known as "Aerosporin," was early shown to be a basic substance forming salts with mineral acids and insoluble precipitates with acidic dyestuffs. The observation of Catch⁴ that the purest preparations gave the ninhydrin reaction in solution, both before and after acid hydrolysis, gave reason to suppose that part of the molecule, at least, consisted of polypeptide. The application of the method of partition chromatography on paper (Consden, Gordon, and Martin⁵) to the products of the acid hydrolysis of polymyxin A preparations by Jones⁶ led to the discovery of the presence of leucine and threonine and of a basic substance with some of the properties of an amino acid, as major constituents. The basic substance was later identified with α,γ -diaminobutyric acid and isolated from the mixed amino acids of the hydrolysate by precipitation with picric acid, as the *DL*-picrate (Catch and Jones⁷). This was the first time that this amino acid, known from the synthetic work of Karrer *et al.*,⁸ Synge,⁹ and Adamson,¹⁰ was identified as a constituent of a natural product. This acid has since been shown to be present in hydrolysates prepared under milder conditions as the *L*-form.¹¹ Receipt of a sample of polymyxin D, the 'polymyxin' of Stansly *et al.*,³ kindly made available by Dr. Carey of the Lederle Laboratories Division of the American Cyanamid Company, enabled us to identify the same amino acids in this antibiotic, with serine as an additional acid. The optical configuration of the leucine and threonine present in polymyxin A has been determined on the paper chromatogram by an enzymic method (Jones¹²). These amino acids were shown to be *D*-leucine and *L*-threonine. Subsequent isolation of *D*-leucine and *L*-threonine from the hydrolytic products by Catch, Jones, and Wilkinson¹¹ has amply confirmed the result obtained by the micro-technique.

It was noted early that, when adequate samples of relatively crude polymyxin A were hydrolyzed by acid in sealed tubes, globules of an oily appearance separated from the acid solution and a rancid odor was present. Evaporation of the acid on a steam-bath in an open dish left only a crystalline residue. The fatty material was at first discounted as an impurity, but was noted for further study. A private communication from Dr. Malcolm of the Lederle Laboratories Division of the American Cyanamid Company informed us of the presence, in the polymyxin molecule, of a

carboxylic acid similar to pelargonic acid. Examination of our purest preparations of polymyxin A showed that there was little ether-soluble material present in the intact substance but that, after acid hydrolysis, ether indeed extracted an acidic substance. Removal of the ether and fractional distillation *in vacuo* yielded an optically active mobile oil, having the properties and composition of a saturated fatty acid $C_9H_{18}O_2$, different from pelargonic acid but as yet unidentified, a description of which is given in the accompanying paper.¹¹

Partition Chromatography of Polymyxin A and Polymyxin D. The elegant method of Consden, Gordon, and Martin for the partition chromatography on filter paper of amino acids and lower peptides, has already been mentioned in describing the discovery of the amino-acid components of polymyxin A. In this method, filter paper, of which the moisture content is maintained constant, becomes the support for this stationary aqueous phase and a mobile, usually largely immiscible, phase is caused to flow over the sheet. The mixture, applied to the paper in solution in a small drop of water or other volatile solvent, which is then removed, is thus subjected, during the flow of the mobile phase, to countless distributions between the two phases, and the components separate as spots. The positions taken up on the paper are dependent mainly upon the distribution coefficients of the components of the mixture for the two phases employed.

The method has the two-fold value not only of making possible the separation and identification of complex mixtures but of providing constants, mainly dependent upon the solubility relationships of substances being studied, with the consumption of a minimum of material.

The almost unparalleled success of this method, which involves the simplest of apparatus, when applied by a number of workers to a variety of classes of organic and inorganic compounds, was an encouragement to attempt the chromatography of polymyxin A.

Given the possibility of separating the components of the admittedly impure preparations of the antibiotic, it was hoped to provide the means by which polymyxin A and its impurities could be recognized.

The attempt was successful and led to the first description of the application of this method to substances of much higher molecular weight than hitherto attempted.⁵ Of the variety of solvents available for the mobile phase, that mixture of butanol, acetic acid, and water recommended by Partridge¹² proved the most suitable. Polymyxin A was applied to the paper in solution in water, 5 μ l. of a solution containing 100 μ g. of the hydrochloride were usually pipetted directly on to a predetermined mark on the paper and the spot allowed to dry. After running the solvent for a suitable time in an atmosphere saturated with respect to all the solvents involved, the antibiotic was found to have moved some 5 to 10 cm. in the direction of the solvent flow, and its position was revealed by fluorescence in ultraviolet light on the dried paper or by the ninhydrin reaction. Preparations of polymyxin A of adequate purity moved as single, rather elongated, spots, but the reproducibility of the positions of the spots on the

paper, relative to the solvent front, in different chromatograms, prepared under what appear to be identical conditions, is not as good as that of simple amino-acids.

A typical chromatogram obtained with polymyxin A is shown in FIGURE 1, section a. The solvent flow is in the direction of the arrow, with the solvent front completely off the leading edge of the paper. The receipt of polymyxin D led to a comparison which showed that polymyxin A and polymyxin D moved on the paper at different rates. FIGURE 1, section c, shows the chromatogram of polymyxin D. Section b, in the same figure, is the chromatogram derived from a mixture of the two antibiotics, using the same amounts as in sections a and c, and, within the limits of the method, this chromatogram is the superimposition of those for the single antibiotics. There can be no doubt that the two antibiotics are different substances. These observations proved to be of particular value at the time of the receipt of the polymyxin D sample since this was known from its stated assay value to be at most 67.5 per cent pure, based on the published figure for the purest material, and the amino acid serine, additional to the amino acids known to be present in polymyxin A, might otherwise have been considered to be a component of an impurity.

Antibiotics from Further Strains of B. AEROSPORUS. The two antibiotics considered above, while exhibiting certain differences, were shown by Brownlee, Bushby, and Short¹⁴ to have some biological properties in common. The important property of renal toxicity (Brownlee and Bushby¹⁵) was one common feature, and, chemical purification having failed to remove the toxic principle, a search by strain selection for less damaging material was instituted in these Laboratories. Examination of the products of fifteen other strains has led to the recognition of other antibiotics, some characterized by their chromatographic behavior as the intact material and others needing the more detailed analysis following hydrolysis. At the time of writing, five groups are known,^{*} including polymyxin D supplied by the American Cyanamid Company.

FIGURE 2 is a reproduction of a chromatogram prepared from only seven of the strains tested. The products of *B. aerosporus* (*B. polymyxa*) are identified here by the Wellcome Foundation Culture Numbers of the selected strains, and the polymyxin D chromatograms were derived from three separate samples from strain B-71 of the Lederle Laboratories. In this figure, three separate groups are recognizable. Strain CN1419 produces an antibiotic which moves fastest on the paper. The strains CN1984, CN2136, CN114J, CN2002, and CN121 yield products which move at comparable speed and more slowly than does CN1419. Polymyxin D moves at an intermediate rate. Further reference to FIGURE 1 shows that the materials derived from strain CN1419 (section d) and polymyxin D (section f) indeed move at different rates and that when mixed (section e),

^{*} The fifth member of the series mentioned previously has, by agreement, been named polymyxin E.¹⁷ Its qualitative composition is identical with that of polymyxin A, but its speed on the paper is approximately equal to that of polymyxin B.

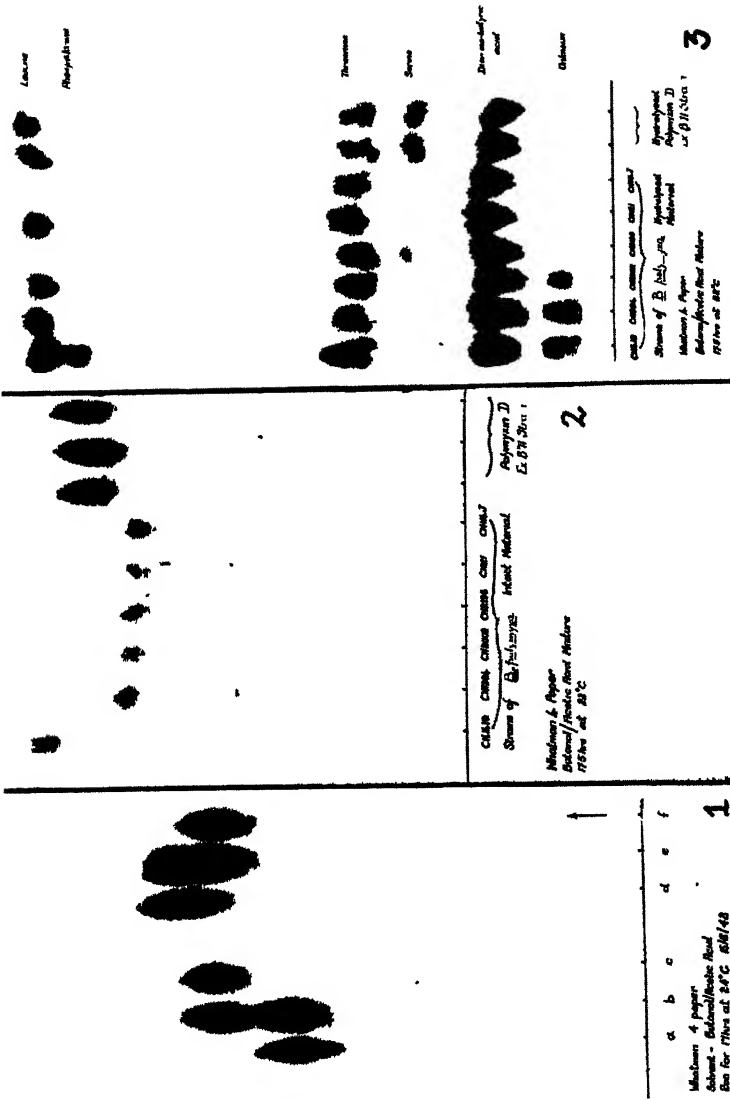


FIGURE 1. Chromatograms of polymyxins a—polymyxin A, b—polymyxins A and D mixed, c—polymyxin D, d—polymyxin B, e—polymyxin B and D mixed, f—polymyxin D.

FIGURE 2. Chromatograms of polymyxins produced by different strains of *B. polymyxa*. Intact material.

FIGURE 3. Chromatograms of polymyxins produced by different strains of *B. polymyxa*. Hydrolyzed materials.

a partial separation may be achieved. The distinction between the products of these strains is of importance and has been confirmed using other solvents. If chromatograms derived from intact material provided the only evidence of the chemical identification of these antibiotics, only three groups would be recognizable from the diagram. If one bears in mind that the speed of migration on the paper is a consequence of a single physical property—the relative distribution between the phases employed—it is clear that a closer analysis might reveal the underlying reason for the different rates and also yield further differences which are not reflected in the similarity of overall distribution coefficients. Such an analysis is provided in the next section.

Partition Chromatograms of the Hydrolyzed Antibiotics. The hydrolytic preparation of the antibiotics for chromatographic analysis was usually performed as follows: Ten mg. of the antibiotic dissolved in 1 ml. 5N HCl

TABLE 1
THE AMINO-ACID COMPONENTS OF THE POLYMYXINS

Polymyxin	Culture No.	Leucine	Phenyl- alanine	Threonine	Serine	α,γ -Diaminobutyric acid
-A	1984 2002 121	+	-	+	-	+
-B	1419	+	+	+	-	+
-C	2136 114J	-	+	+	-	+
-D	B71 "Lederle"	+	-	+	+	+

was placed in a sealed tube and kept at 100°C. for 2.5 hours. The HCl was then evaporated on an open dish and the residue dehydrated by evaporation with alcohol and benzene. The residue, which crystallized readily, was dissolved in 0.25 ml. of water and 5 μ l. of this solution, equivalent to 200 μ g. of the mixed amino acids, was placed on the paper for the chromatogram. This relatively large quantity of amino acid mixture was necessary because of the disproportionately small amount of some amino acids present.

The results of the examination of the seven strains, shown for the intact material in FIGURE 2, are given in FIGURE 3, which depicts a chromatogram formed on Whatman 4 paper, using the butanol/acetic acid/water mixture previously mentioned. The main amino acid spots are readily identified as leucine, phenylalanine, threonine, serine, and α,γ -diaminobutyric acid, in that order downwards. The direction of flow of the solvent on the paper is from the line at the bottom in an upward direction. One-dimensional partition chromatography sufficed to separate the small number of amino acids present.

We now have the data for the classification of the products of the *B. polymyxa* group of organisms, and this is shown in TABLE 1. Inspection of FIGURE 3 shows that, in agreement with FIGURE 2 and as an extension of the findings based on that figure, there now appear to be four distinct classes of antibiotics, the component amino acids of which are shown in TABLE 1. It was evident that a new nomenclature was necessary for the identification of the classes revealed by these observations, and, for reasons which are given by Brownlee¹⁶ in the introductory paper of this series, the name, polymyxin, was agreed upon as the generic name. The classes of antibiotic are then distinguished by alphabetical suffixes, such as appear earlier in this paper. TABLE 1 provides the evidence upon which the present assignment of names is based. This evidence is thus confined, at present, to the qualitative amino acid composition of the antibiotics. The basis of the classification may require modification at a later date, should evidence of quantitative differences distinguish new members of the group. In TABLE 1, the signs given under the headings of the amino acids indicate spots of an intensity known by experience to represent a significant amount of the amino acid. The spots given by phenylalanine in the reproduction of FIGURE 3 are much fainter than those of the freshly-prepared chromatogram. The color of the spot following use of the acid solvent mentioned previously is characteristic for phenylalanine. Other relatively faint spots are present in the chromatogram, e.g., CN2136 and CN114J give products in which there is a faint spot in the leucine position, and spots in the serine position appear in the chromatograms given by other products. The question whether these classes of polymyxin represent single antibiotics or mixtures of two or more cannot be answered in our present state of knowledge. Most of the materials investigated in this comparative study are not more than 50 per cent pure. The fainter spots could, therefore, be due to the presence either of impurities devoid of antibacterial activity, or of small proportions of other antibiotics. To give an example, the faint spots for leucine in the polymyxin C chromatograms could be due to admixture of a small proportion of polymyxin A. That polymyxin B is not a mixture of polymyxins A and C, which could be one interpretation of its amino acid composition, is shown, chemically, by the chromatograms of the intact materials (FIGURE 2), and, biologically, by the absence of nephrotoxicity¹⁴ even in preparations known to be impure. Should purification of the polymyxin C preparations yield homogeneous material still showing the presence of leucine in the smaller proportions indicated before, the molecular weight of this antibiotic would be rather greater than the present estimates based on polymyxin D and polymyxin A. What is not known at present is whether each class is a single antibiotic of settled structure or a mixture of several substances containing the same amino acids. The available evidence based on the purification of polymyxin A¹¹ and on the present work appears to favor the concept that each strain of the organisms produces one main antibiotic, which, when isolated in a standard manner, appears substantially homogeneous.

An Important Impurity. An impurity which is present in many prep-

arations of polymyxin A, B and C, but not in polymyxin D, and which requires multiple purification procedures for its removal, is usually observed as a faint, slow spot near the point of application. In the chromatograms of the hydrolysis products (FIGURE 3), there appears a series of spots of all grades of intensity, corresponding with the label "unknown." These latter spots are thought to be related to the former and their position on the chromatograms, both of the intact and hydrolysed material, suggests that they are very basic substances. The material is described in greater detail in the accompanying paper.¹¹

The Fatty Acid Component. As mentioned above, the fatty acid components of these preparations have been characterized but not yet identi-

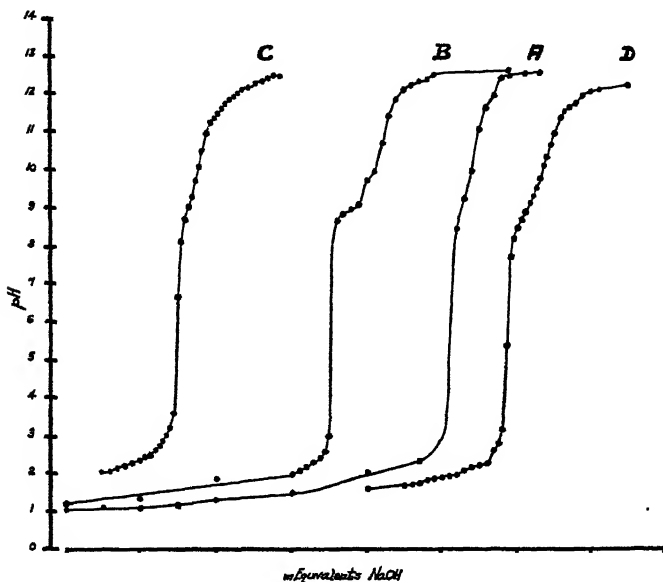


FIGURE 4. The titration curves of polymyxins A, B, C and D. The curves begin at arbitrary values of the abscissa for clarity of presentation.

fied, and the acids have been isolated from polymyxin A, B, and D as the *p*-bromobenzylthiuronium salts. These show identical crystalline form, melting points, and mixed melting points.¹¹

Titration Curves. The electrometric titration curves for the purest representatives of the four classes of polymyxin are shown in FIGURE 4. The solution of the antibiotic in water (0.2 per cent solution) was titrated with 5N HCl to pH 1-2 and then titrated to pH 12 with 2N NaOH using a calibrated glass electrode. The curves show the behavior to be expected from relatively strongly basic substances but there are two additional inflexions of different extent. The inflexions are not equal for the four preparations, and the most alkaline inflexion, while believed to be real, is only barely re-

moved in extent from the limits of experimental error. The inflexion in the region pH 9 to 10 is much more pronounced in the case of polymyxin B. If this corresponded with the presence in the molecule of free carboxyl groups, it would indicate a lower isoelectric point for this preparation. This is a fact which is known from other data.

Ultraviolet Spectra. The ultraviolet absorption spectra of polymyxin B and C, as would be expected, show the peak corresponding with phenylalanine but general absorption prevents the use of this for quantitative purposes.

Summary

1. The application of paper chromatography to the antibiotics prepared from different strains of *B. polymyxa* has revealed the multiple nature of this antibiotic group.

2. The polymyxins are classified, partly by the rate of movement of the intact material on the paper relative to the solvent front, and partly by their qualitative amino acid composition.

3. The following acids have been found to enter into the composition of all polymyxins studied; threonine, α,γ -diaminobutyric acid, and an unidentified, optically active, fatty acid of empirical formula $C_9H_{18}O_2$.

4. Polymyxin A has D-leucine, and polymyxin C, phenylalanine, additional to the above. Polymyxin B has both leucine and phenylalanine. Polymyxin D has the additional amino acids leucine and serine.

5. Single strains of the organism appear to produce single antibiotics.

6. The titration curves of representatives of the classes mentioned show similarities and differences. Interpretation is reserved, except that polymyxin B is shown to have a lower isoelectric point than the others.

Bibliography

1. AINSWORTH, G. C., A. M. BROWN & G. BROWNLEE. 1947. *Nature* **160**: 263.
2. BENEDICT, R. G. & A. F. LANGLYKKE. 1947. *J. Bact.* **54**: 24.
3. STANSLY, P. G., R. G. SHEPHERD, & H. J. WHITE. 1947. *Bull. Johns Hopkins Hosp.* **81**: 43.
4. CATCH, J. R. & R. FRIEDMANN. 1948. *Biochem. J.* **42**: lii.
5. CONSDEN, R., A. H. GORDON, & A. J. P. MARTIN. 1944. *Biochem. J.* **38**: 224.
6. JONES, T. S. G. 1948. *Biochem. J.* **42**: xxxv.
7. CATCH, J. R. & T. S. G. JONES. 1948. *Biochem. J.* **42**: lii.
8. KARRER, P., K. ESCHER, & R. WIDMER. 1926. *Helv. Chim. Acta* **9**: 301.
9. SYNGE, R. L. M. 1939. *Biochem. J.* **33**: 671.
10. ADAMSON, D. W. 1939. *J. Chem. Soc.* 1564.
11. CATCH, J. R., T. S. G. JONES, & S. WILKINSON. 1948. *Ann. N. Y. Acad. Sci.* **51** (5): 917-923.
12. JONES, T. S. G. 1948. *Biochem. J.* **42**: lix.
13. PARTRIDGE, S. M. 1948. *Biochem. J.* **42**: 238.
14. BROWNLEE, G., S. R. M. BUSHBY, & E. I. Short. 1948. *Ann. N. Y. Acad. Sci.* **51** (5): 891-896.
15. BROWNLEE, G. & S. R. M. BUSHBY. 1948. *Lancet* **1**: 127.
16. BROWNLEE, G. 1948. *Ann. N. Y. Acad. Sci.* **51**(5): 875-878.
17. JONES, T. S. G. 1948. *Biochem. J.* **43**: xxvi.

THE CHEMISTRY OF POLYMYXIN A

By J. R. CATCH, TUDOR S. G. JONES, AND S. WILKINSON*

The Wellcome Chemical Research Laboratories, Beckenham, Kent, England

The object of this paper is to describe the present state of our knowledge of polymyxin A, which is the agreed name for the antibiotic originally called "Aerosporin" by Ainsworth, Brown, and Brownlee.¹ At the same time, where data are available, we shall compare this antibiotic with the related antibiotics polymyxin B, discovered in our laboratories, and polymyxin D, the "polymyxin" of Stansly *et al.*² Experimental details will be published elsewhere and only the findings based on them will be given here.

Polymyxin A

Isolation, Purification, and Properties. The isolation and partial purification of polymyxin A has been described by Catch and Friedmann³ in abstract form and will be repeated here in somewhat greater detail to provide data for the materials used in the subsequent work to be described. Crude polymyxin A, prepared according to Ainsworth, Brown, and Brownlee,¹ is purified by fractional precipitation as helianthate, successive portions of helianthic acid in dilute aqueous pyridine being added to the neutral antibiotic in water or aqueous methanol. The polymyxin A is generally precipitated first, followed by impurities which appear to be, at least in part, basic peptides. The helianthate is an amorphous brown powder, almost completely insoluble in water and most solvents. The hydrochloride or sulfate is recovered by treatment of a methanolic suspension with the appropriate mineral acid, removal of the sparingly soluble helianthic acid, and precipitation with anhydrous acetone.

FIGURE 1 shows the paper partition chromatogram prepared by methods described in the preceding paper⁴ from equal quantities by weight of the hydrochlorides obtained from six successive partial precipitations of a single batch. It should be emphasized that the earlier fractions were greater both in quantity and activity than the later ones. A basic impurity is seen to appear in the third fraction and to increase in amount as judged by intensity as the series progresses. No separation of the nephrotoxic principle (Brownlee and Bushby⁵) was observed to have taken place, since the six fractions, when administered parenterally to rats on a unitage basis, were all equally toxic.⁶

The further precipitation of the amorphous reineckate and regeneration by Ag_2SO_4 and BaCl_2 , or better, by suspending in dilute hydrochloric acid and extracting the free reinecke acid with *cyclohexane*, yields a highly active hydrochloride in over 99 per cent yield based on activity.

FIGURE 2 shows the chromatogram of the hydrochlorides produced by

* Acknowledgment is gratefully made to Dr. Kathleen Lonsdale, F.R.S., for her provision of X-ray powder photographs of the amino acids.

fluorescence in the ultra-violet. Sulfur and phosphorus are absent. The elementary composition of a representative hydrochloride is

C, 45.5; H, 7.7; N, 15.8; Cl, 14.0%.

Many base precipitants and acid dyes give precipitates with solutions of the antibiotic. The picrate and flavianate are amorphous and gummy. Attempts to prepare salts of organic acids by treatment of the hydrochlorides with the silver salts of acetic, phenylacetic, oxalic, malic, benzoic, phthalic, cinnamic, succinic, and naphthalene- β -sulphonic acids yielded in all cases amorphous and sometimes glassy materials.

Amino Acid Components. The amino acid components of polymyxin A have been shown by Jones⁷ and by Catch and Jones⁸ to be leucine, threonine, and α, γ -diaminobutyric acid. The former author has shown,⁹ by application of the method of Synge,¹⁰ which employs a *D*-amino-acid oxidase¹¹ for the identification of certain *D*-amino-acids on paper chromatograms, that the optical configurations correspond to *D*-leucine and *L*-threonine. The *D*-amino-acid oxidase oxidizes *D*- α, γ -diaminobutyric acid slowly (*cf.* ¹²) and no evidence could be obtained for the configuration of this component. Catch and Jones⁸ described the isolation of *DL*- α, γ -diaminobutyric acid from the products of the hydrolysis of polymyxin A, *via* the dipicrate dihydrate, m.p. 185°, as the monohydrochloride, m.p. 228–230° dec. This picrate, although readily obtained from the residue left after evaporation of the hydrolysis mixture, was contaminated by the picrate of at least one other base, and serial fractional crystallization of the derived, mixed hydrochlorides was necessary before the pure *DL*- α, γ -diaminobutyric acid was obtained. Moreover, the precipitation of the diamino-acid was incomplete, as judged by paper chromatograms, and fractional crystallization from alcohol, of the residue, after removal of excess picric acid and evaporation, yielded a small amount of *L*-diaminobutyric acid. From the residual mother liquors, small quantities of impure *D*-leucine and *L*-threonine were obtained. Subsequent work has shown that the conditions employed in the early experiments, namely, hydrolysis at 150°C. for 2.5 hours with a concentration of 5N hydrochloric acid, were unnecessarily drastic, and that a temperature of 100°C. sufficed. Polymyxin A, which had been hydrolyzed under the latter conditions, has yielded mainly the *L*-form of the diamino-acid with a little of the *DL*-form. From the same hydrolysate, *L*-threonine and *D*-leucine have also been obtained in the pure form.

The precipitation with picric acid has been found to be unnecessary, since the direct fractional crystallization from alcohol of the residue obtained by evaporation of the ether-extracted acid hydrolysate has rendered possible the easy separation of the diamino-acid in the pure state. In this way, both the *L*- and the *DL*-forms have been isolated in the ratio 3:1. The occurrence of *DL*- α, γ -diaminobutyric acid in the hydrolysate raises the question of the presence of the *D*-form in the intact antibiotic. It is known from control experiments that the treatment of the *L*-form with

5N hydrochloric acid at elevated temperatures leads to some racemization. This is thought to be insufficient to account for the proportion of the *DL*-form isolated, and it is considered possible that racemization at the moment of liberation of the *L*-diamino-acid, rather than during the actual course of the hydrolysis, is the source of the *D*-form.

The isolation of the neutral amino acids was accomplished by the following methods: fractional precipitation with pyridine of the mother liquors from the crystallization of the diamino-acid yielded *L*-threonine, and *D*-leucine was precipitated from the filtrate by naphthalene- β -sulphonic acid. The isolation of the amino acids thus confirms the determination of the optical configurations of the neutral amino acids by means of *D*-amino-acid oxidase,⁸ and extends the characterization to the diamino-acid, for which no evidence for optical form could be obtained by the enzymic

TABLE 1
ANALYTICAL DATA AND YIELDS OF THE AMINO ACID COMPONENTS OF POLYMYXIN A

Component	C	H	N	Cl	$[\alpha]_{5461}^{21}$	Yield %
<i>L</i> - α , γ -Diaminobutyric acid monohydrochloride: Found	—	—	17.9	22.9	+28° ($c = 0.96$, 6N HCl)	46
$C_4H_{11}N_2O_3Cl_2$ requires	—	—	18.1	23.0		
<i>DL</i> - α , γ -Diaminobutyric acid monohydrochloride: Found	31.3	7.0	18.1	22.7	0° ($c = 1.11$, 6N HCl)	15
$C_4H_{11}N_2O_3Cl_2$ requires	31.1	7.2	18.1	22.9		
<i>DL</i> - α , γ -Diaminobutyric acid dipicrate: Found	33.7	2.6	19.5	—	—	—
$C_{16}H_{18}N_8O_{16}$ requires	33.3	2.8	19.5	—	—	—
<i>L</i> -threonine: Found	40.5	7.8	11.8	—	-32.6° ($c = 0.2$, water)	6
$C_6H_9NO_3$ requires	40.3	7.6	11.8	—		
<i>D</i> -leucine: Found	54.8	10.2	10.9	—	-15° ($c = 0.13$, 6N HCl)	3
$C_6H_{13}NO_3$ requires	54.9	10.0	10.6	—		

method. The identity of the three amino acids with the synthetic, optically active acids has also been confirmed, through the courtesy of Dr. Kathleen Lonsdale of University College, London, by X-ray powder photographs of the crystalline material. The analytical data together with the yields of the amino acids isolated from one batch of polymyxin A, of approximately 50 per cent purity, are shown in TABLE 1.

During the course of the isolation of the products of the hydrolysis, paper partition chromatography has proved of immense value, both for identifying rapidly the small quantities of crystalline material which separated, and also for following the purification of these substances. By suitable choice of the quantity applied to the paper, it has been possible to demonstrate the presence of impurities which escaped detection by the ordinary criteria of melting point, elementary analysis or specific optical rotation.

An Important Impurity. The basic impurity mentioned above, the presence of which in polymyxin A is demonstrated in FIGURE 1, has been noted to give a more intense blue reaction with ninhydrin, both on paper and in solution, following acid hydrolysis. From the products of hydrolysis of the later fractions shown in FIGURE 1, a picrate melting at 252° after softening at 250° was readily obtained, and this yielded a hydrochloride, m.p. 212–213°. Analyses of the picrate and the hydrochloride are consistent with those for derivatives of a compound of formula $C_5H_{14}N_2$ as is shown below:

	C%	H%	N%	Cl%
Hydrochloride: Found	—	—	16.0	40.6
$C_5H_{14}N_2Cl_2$ requires	—	—	16.1	40.8
Picrate: Found	36.2	3.6	20.3	—
$C_5H_{14}N_2 \cdot 2C_6H_5N_2O_7$ requires	36.4	3.6	20.0	—

TABLE 2
ANALYTICAL DATA FOR THE FATTY ACID DERIVATIVES

Component	C	H	N	Br
<i>p</i> -Bromobenzylthiuronium salt: Found	50.5	6.6	6.6	20.2
$C_{17}H_{28}N_2O_2SBr$ requires	50.6	6.7	7.0	19.9
amide: Found	68.7	11.8	8.8	—
$C_9H_{19}NO$ requires	68.8	12.1	8.9	—
methyl ester: Found	69.4	11.7	—	—
$C_{10}H_{21}O_2$ requires	69.8	11.6	—	—

The hydrochloride and picrate are optically inactive. The melting points of mixtures of the hydrochloride and picrate with the corresponding salts of cadaverine are depressed. The substance remains unidentified.

The Fatty Acid Component. The presence of oily material, of a rancid odor, in the products of the hydrolysis of relatively crude polymyxin A was noted and the separated fatty material reserved for later study. A private communication from Dr. Malcolm of the Lederle Laboratories Division of the American Cyanamid Company, stating that polymyxin D contained about 15 per cent of an acid similar to pelargonic acid, led us to investigate this fatty material in greater detail. Ether extracted little material from the polymyxin A preparation used, but, after acid hydrolysis, 8–10 per cent by weight of the antibiotic was obtainable as a crude oil, and repeated distillation *in vacuo* yielded 5.5 to 7.5 per cent of a pure, mobile oil, from which the *p*-bromobenzylthiuronium salt, m.p. 160–161°, was obtained. This was reconverted to the acid and the latter yielded a crystalline amide, m.p. 93°C. The fatty acid was optically active, $[\alpha]_{D}^{25} = +8.6^\circ$ ($c = 2.22$, ether) and yielded an optically active methyl ester, b.p. 35–36°C./0.01 mm. Hg, $[\alpha]_{D}^{25} = +9.0^\circ$ ($c = 1.55$, ether). The analytical data for the salt, the amide, and the ester are shown in TABLE 2 and are seen to agree with those of an acid of empirical formula $C_9H_{18}O_2$.

The treatment of polymyxin A with 2N alkali at room temperature led to complete destruction of the optical and biological activity in 7 hours. No fatty acid was extractable from the end product, but further *acid* hydrolysis liberated a fatty acid, which was isolated and gave a *p*-bromobenzylthiuronium salt identical with the one already described. This observation suggests that the fatty acid radical cannot, in all probability, be present in the molecule of polymyxin A as an *O*-acyl group, since the comparable compounds, *O*-acetyl-hydroxyproline and *O*-acetyl-threonine,

TABLE 3
YIELD OF FATTY ACIDS

Source	u./mg.	Per cent crude acid	Per cent redistilled fatty acid
Polymyxin A*	5360	8	5.5
	6150	7.8	5.6
	7940	8	7.2
	9000	9.7	7.5
Polymyxin B*	4800	6.3	5.4
Polymyxin D†	1350	12.2	—

* Wellcome Foundation 1947 Standard.

† Cyanamid units.

TABLE 4
PROPERTIES OF THE ACIDS ISOLATED

Source	Refractive index	m.p. of <i>p</i> -Bromobenzylthiuronium salt		Mixed melting point			m.p. of Amide	
		°C.	Inserted at	Mixture	m.p.	Inserted at	m.p. °C.	Mixed m.p.
Polymyxin A	1.4390 (18°C.)	159	155	A & B	159-160	155	93	93
Polymyxin B	1.4330 (20°C.)	160	155				93	
Polymyxin D	—	158-158.5	150	B & D	158.9	155		

are easily hydrolysed by cold alkali. If the fatty acid were present as the anion of one or more of the basic amino-groups, then extraction with ether of the acid solution of intact polymyxin A should yield the fatty acid. It is probably present as an *N*-acyl grouping.*

Polymyxin B and Polymyxin D

Amino Acid Components. The existence of other antibiotics of the polymyxin class has been demonstrated in the preceding paper.⁴ The

* Subject to final confirmation by synthesis, the constitution of the fatty acid has been established by degradative studies as 6-methyl-octan-1-oic acid.



amino acids present in these antibiotics are given in TABLE 1 of that paper. The optical configurations of these amino acids have yet to be determined, with the exception that *L*- α , γ -diaminobutyric acid has been isolated from polymyxin B by direct fractional crystallization from alcohol of the evaporated residue from the acid hydrolysis, as described above for polymyxin A. Experiments are now in progress for the determination of the configurations of the other amino acids.

Fatty Acid Component. The isolation of the fatty acids from the acid hydrolysates of polymyxin B and polymyxin D and the preparation of suitable derivatives, has shown that the acids are identical with one another and with that from polymyxin A. The identity is demonstrated in TABLE 3, which gives the melting points and mixed melting points of the *p*-bromobenzylthiuronium salts and of the amides. The yields of the fatty acid from several different samples of polymyxin A and one sample each of polymyxin B and polymyxin D are shown in TABLE 4. It is seen that the sample of polymyxin D gave rather more of the crude fatty acid than did the two others.

Summary

1. The partial purification of polymyxin A (formerly described as "Aerosporin" by Ainsworth, Brown, and Brownlee¹) *via* the helianthate and reineckate is described. The properties of the product are summarized.

2. The products of the acid hydrolysis have been shown to yield *L*- α , γ -diaminobutyric acid, *DL*- α , γ -diaminobutyric acid, *L*-threonine, and *D*-leucine.

3. An optically active fatty acid is also present in acid, but not alkaline hydrolysates. Analyses of its *p*-bromobenzylthiuronium salt, its amide, and its methyl ester agree with the empirical formula $C_9H_{18}O_2$. The acid is not extractable from the unhydrolyzed material by ether, and is therefore unlikely to be an impurity or the anion of one of the basic groups.

4. Polymyxin B has been shown to yield *L*- α , γ -diaminobutyric acid and both polymyxin B and polymyxin D give the same optically active $C_9H_{18}O_2$ acid as does polymyxin A.

Bibliography

1. AINSWORTH, G. C., A. M. BROWN, & G. BROWNLEE. 1947. *Nature* **160**: 263.
2. STANSLY, P. G., R. G. SHEPHERD, & H. J. WHITE. 1947. *Bull. Johns Hopkins Hosp.* **81**: 43.
3. CATCH, J. R. & R. FRIEDMANN. 1948. *Biochem. J.* **42**: lii.
4. JONES, T. S. G. 1948. *Ann. N. Y. Acad. Sci.* **51**(5): 909-916.
5. BROWNLEE, G. & S. R. M. BUSHBY. 1948. *Lancet* (1): 127.
6. BROWNLEE, G., S. R. M. BUSHBY, & E. I. SHORT. 1948. *Ann. N. Y. Acad. Sci.* **51**(5): 891-896.
7. JONES, T. S. G. 1948. *Biochem. J.* **42**: xxxv.
8. CATCH, J. R. & T. S. G. JONES. 1948. *Biochem. J.* **42**: lii.
9. JONES, T. S. G. 1948. *Biochem. J.* **42**: lix.
10. SYNGE, R. L. M. 1948. Private communication; see ref. 9 above.
11. NEGELEIN, E. & H. BRUMEL. 1939. *Biochem. Z.* **300**: 225.
12. HANDLER, P., F. BERNHEIM, & J. R. KLEIN. 1941. *J. Biol. Chem.* **138**: 203.

INFRARED STUDIES

By R. C. GORE AND E. M. PETERSEN

Stanford Research Laboratories, American Cyanamid Company, Stamford, Connecticut

Introduction

Inasmuch as the knowledge gained from spectrometric observations in the infrared region has been of value in the study of the molecular structure and identification of other antibiotics,¹ it was natural to make such observations in the case of polymyxin.

As an introduction to the use of this, as yet, little-used technique in the study of antibiotic molecules it is, perhaps, worth while to summarize the advantages and disadvantages of infrared spectrometric studies. A recent review² gives these in considerably more detail than can be presented here, but they may be summarized as follows:

Advantages

(1) All molecules possess infrared spectra, whether solids, liquids, gases, or in solution.

(2) A single infrared spectrogram may give clues to the identity of constituent radicals, substituents and their positions in a molecule.

(3) The spectrum is a fingerprint of the molecule useful in both qualitative and quantitative analysis.

(4) This spectrogram gives an indication of the overall purity of the preparation.

(5) The techniques are simple.

(6) Relatively small samples of pure materials are sufficient (several milligrams). Samples may be recovered unchanged.

Disadvantages

(1) Although all molecules possess infrared spectra, with present instruments, as the molecular weight increases, the clarity and uniqueness of the spectrum decreases.

(2) The present knowledge of infrared spectra is insufficient to account for and predict all absorption bands. This makes it necessary, as in the case of mixed melting point determinations, to obtain known compounds when identifying an unknown one.

(3) Traces of impurities, unless they are exceedingly strong absorbers, are not easily detected.

(4) Dilute solutions are handled with difficulty, unless strong absorbers are present.

As an indication of the utility of infrared studies, the spectra of five penicillins are shown in FIGURE 1. Unique absorption bands characteristic of each type are indicated at various frequencies in the spectra. Qualitatively, it is possible to point out the absorption bands related to the N—H, carbonyl, and C=C radicals. As is evident, each spectrum is a characteristic physical property of the molecule.

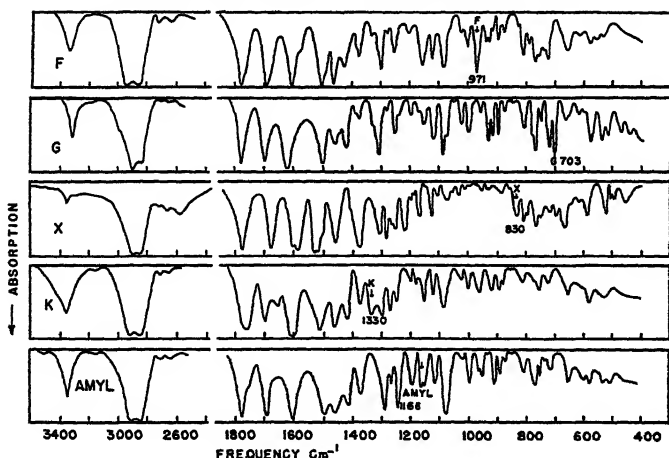


FIGURE 1. Infrared absorption spectra of the crystalline sodium salts of various penicillins.

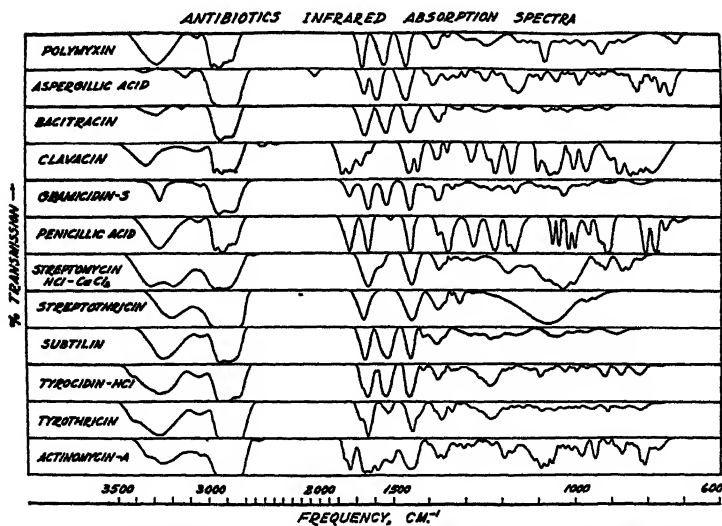


FIGURE 2. Semi-line graph of the infrared absorption spectra of some common antibiotics.

It may be of interest to observe the spectra of several antibiotics taken at random from our catalog of such materials. This file has been assembled through the generosity of many research workers in the field, and especially through the efforts of Dr. W. C. Tobie. At a later date, a more complete story is to be published on the spectra of antibiotics in general. A semi-line-graph of several of these materials is shown in FIGURE 2, again illustrating the uniqueness of the infrared spectra.

Studies of this nature have been and are being made on polymyxin, "Aerosporin," and their breakdown products.

Experimental

Many of the spectra shown have been taken on a Perkin-Elmer spectrometer equipped with electronic direct-current amplification and a Brown recorder. Most of the solid materials were observed by merely mulling the sample with white mineral oil (Nujol) between rocksalt plates. Liquids were observed in rocksalt cells with tin foil spacers. All samples have been furnished by the various workers taking part in this presentation and all credit for their preparation must be given to them.

Results

Spectra of polymyxin, AB-71, and "Aerosporin" are shown in FIGURE 3.

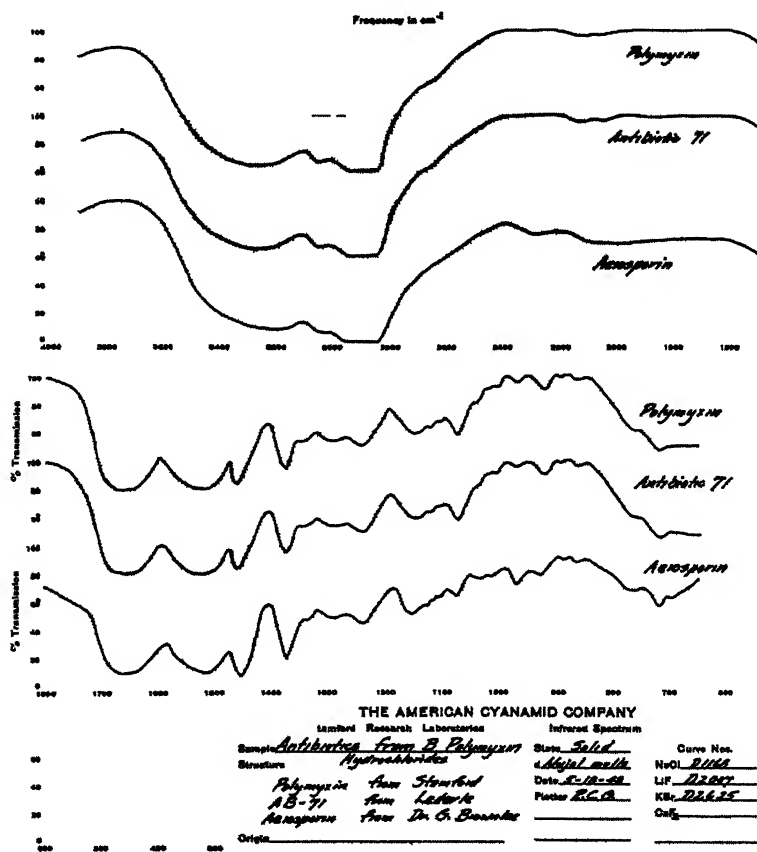


FIGURE 3. Spectra of antibiotics from *B. polymyxa*.

It is evident, by a comparison of these spectra and those of the penicillins, that there are fewer unique, sharp absorption bands. This is a direct conse-

quence of the high molecular weight of the *B. polymyxa* derivatives. A difference in the absorption near 1000 cm^{-1} between "Aerosporin" and polymyxin is real and is indicative of differences in these substances. It is also evident that polymyxin and Antibiotic 71, produced by different methods, are identical.

A C_9 fatty acid from polymyxin, submitted by Dr. S. Kushner, is shown in FIGURE 4. It was thought that this material might be pelargonic acid.

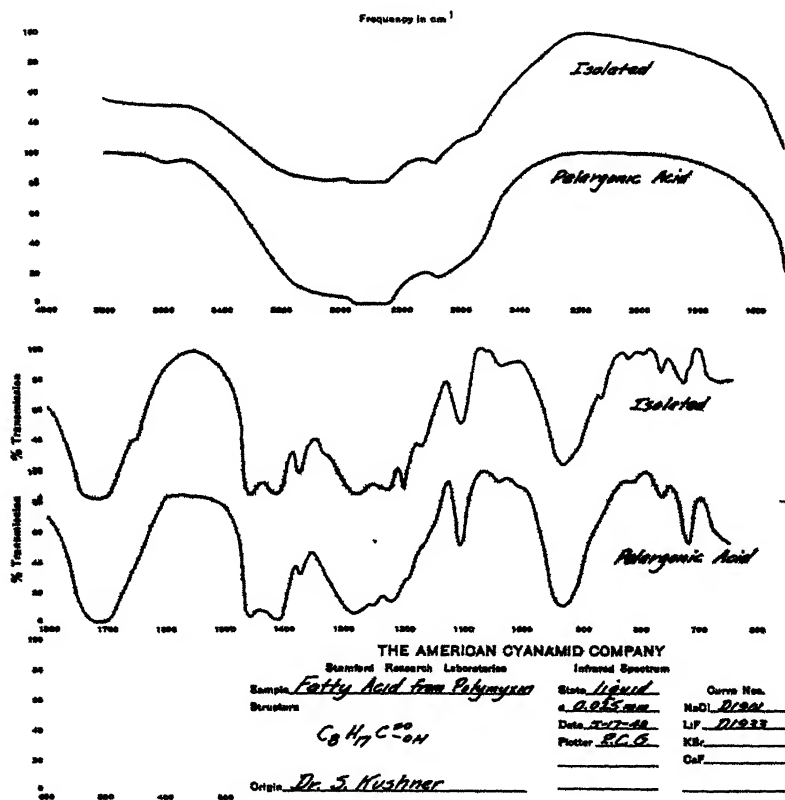
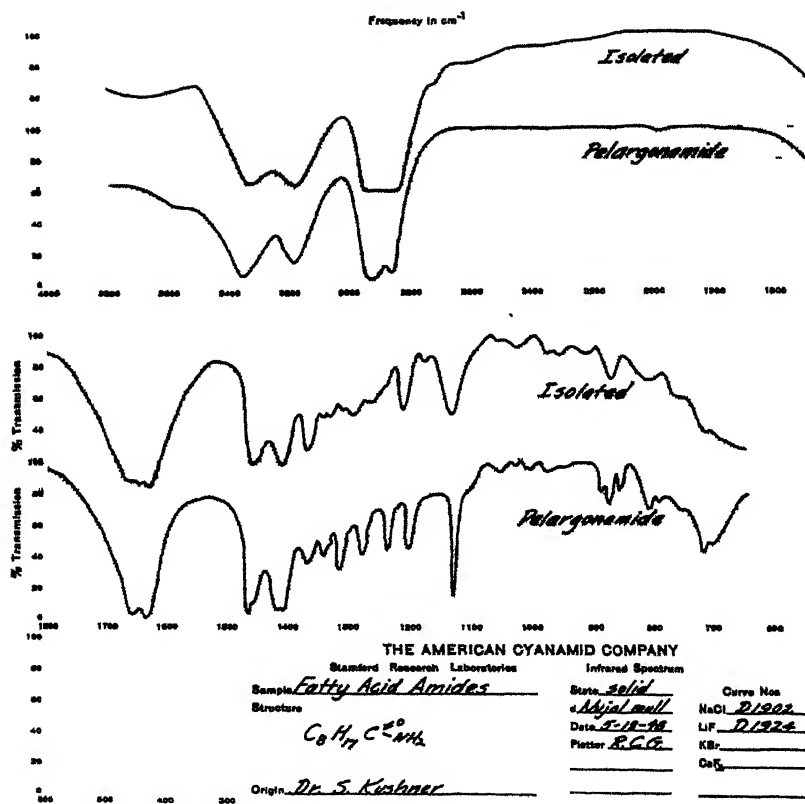


FIGURE 4 spectra of an isolated C_9 fatty acid and pelargonic acid.

A sample of this latter acid, shown for comparison, is similar, but the methyl deformation frequency near 1375 cm^{-1} indicates that the acid is a branched chain, since more absorption is shown in the unknown than in the straight chain pelargonic acid.

As further proof of the lack of identity between the unknown and the pelargonic acid, the spectra of the amides, prepared by Dr. Kushner, are shown in FIGURE 5. The same methyl absorption difference as was shown in the case of the acids is evident, as well as many other differences.

FIGURE 5. Spectra of the amides of the C_{17} fatty acid and of pelargonic acid

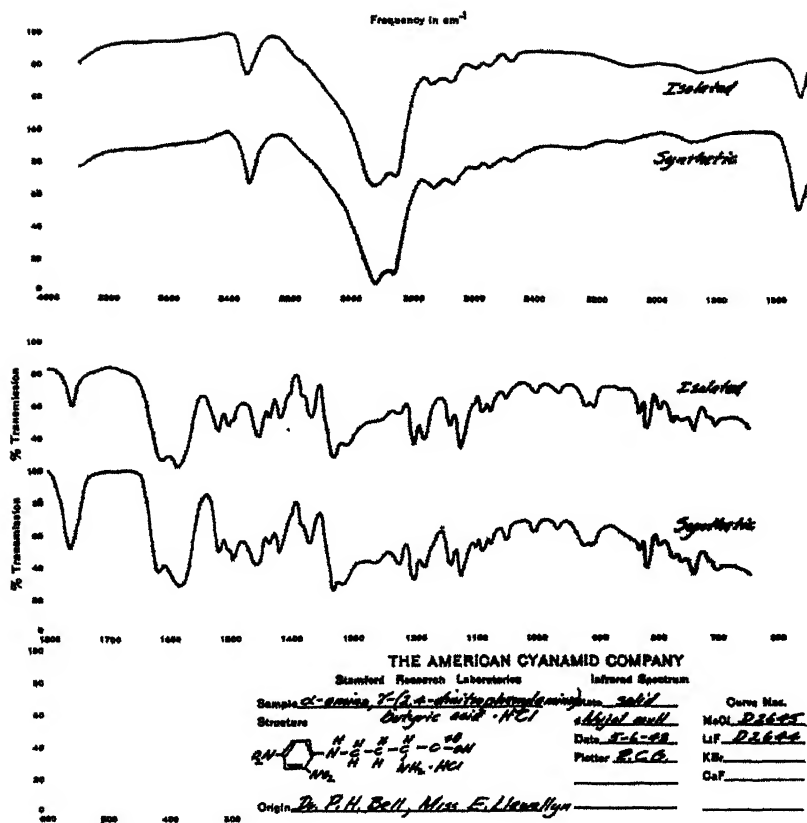


FIGURE 6. Spectra of the isolated and synthetic α -amino γ -(2,4-dinitrophenylamino) butyric acid hydrochlorides.

Samples of the hydantoin of an unknown diamino acid and α, γ -diaminobutyric acid, isolated by Dr. Kushner, were submitted, but because of the possibilities of differences inherent in the preparation of hydantoin the spectra were not conclusive.

The 2,4-dinitrophenyl derivatives of a similar material as well as that of the known acid were submitted by Dr. P. H. Bell. The identity of the materials is indicated by the spectra in FIGURE 6. Similar spectra of the di(2,4-dinitrophenyl) derivatives of the unknown and the *L* (+) α, γ -diaminobutyric acid are shown in FIGURE 7. From a comparison of these

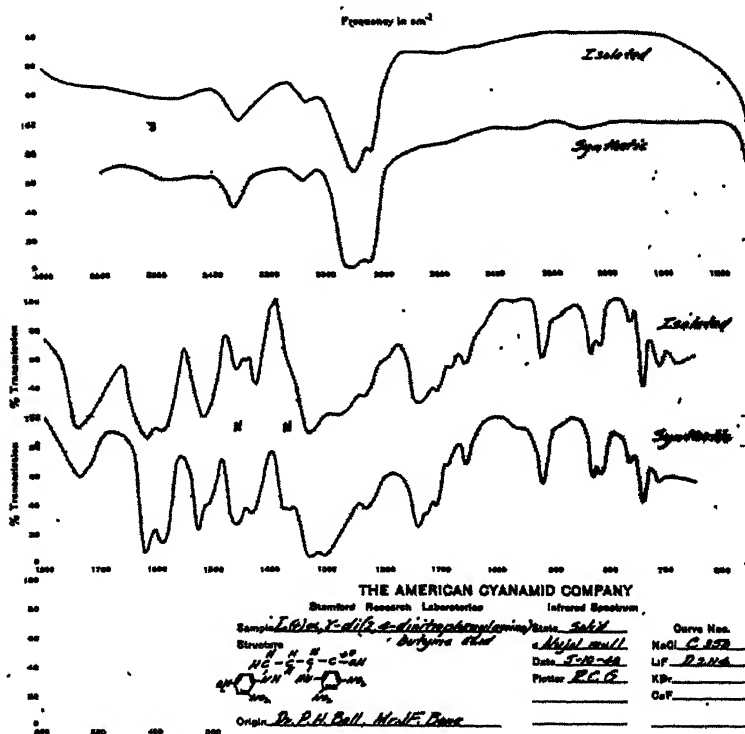


FIGURE 7 Spectra of the isolated and synthetic *L* (+) α, γ -di(2,4-dinitrophenylamino) butyric acids.

spectra, it appears as if the unknown amino acid is *L* (+) α, γ -diaminobutyric acid.

An amino acid, suspected of being serine, was isolated by Dr. P. H. Bell as the 2,4-dinitrophenyl derivative. The spectra of this derivative and that of *D*-serine are shown in FIGURE 8. Although slight purity differences are evident, the material appears to be *D*-serine.

The spectra of an amino acid isolated by Dr. R. G. Shepherd thought to be leucine, along with spectra of *dl*- and *L* (–) leucines are shown in FIGURE 9.

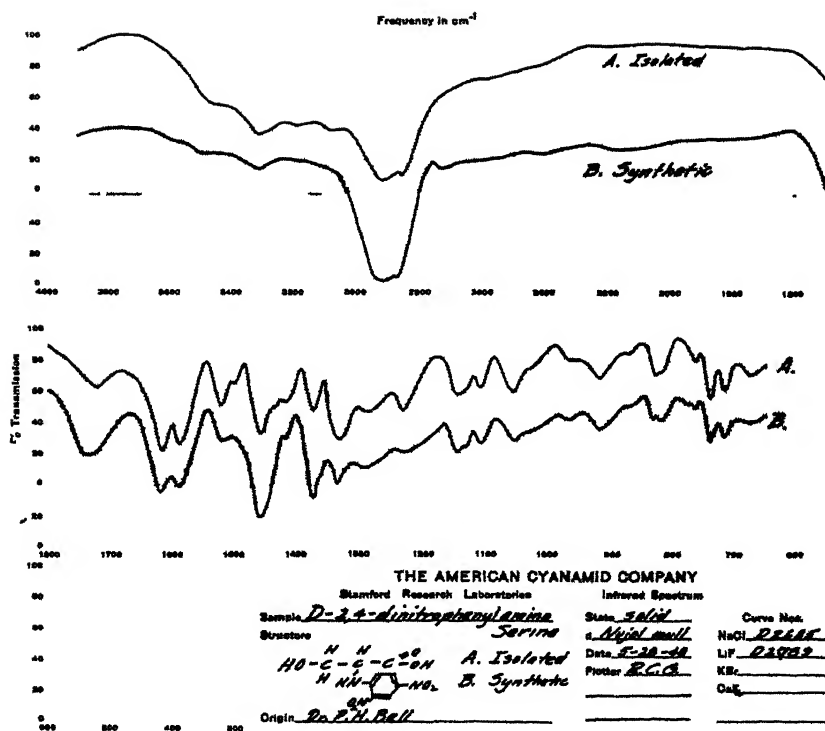
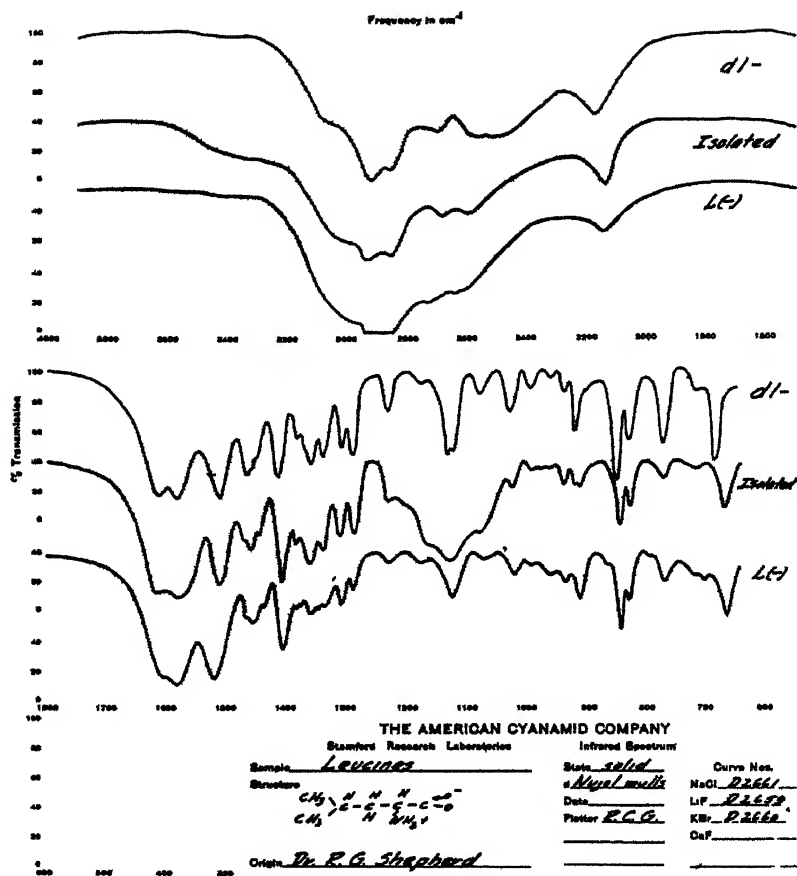


FIGURE 8. Spectra of the isolated serine derivative and synthetic D-2,4-dinitrophenylamino serine

FIGURE 9. Spectra of the *dl*, isolated, and *L* (+) leucine.

It is to be remembered that these spectra were taken in the solid phase where enantiomorphic differences might occur. Although the spectra of the isolated material and the knowns show differences in several regions, especially near 1125 cm^{-1} , it may be concluded that this difference is caused by impurities and the isolated material is *L*(-) leucine because of the correspondence of the absorption bands near 670 cm^{-1} .

Another amino acid, suspected of being threonine, along with the spectra of known *D* (-) and *L* (+) threonine are shown in FIGURE 10. As can be

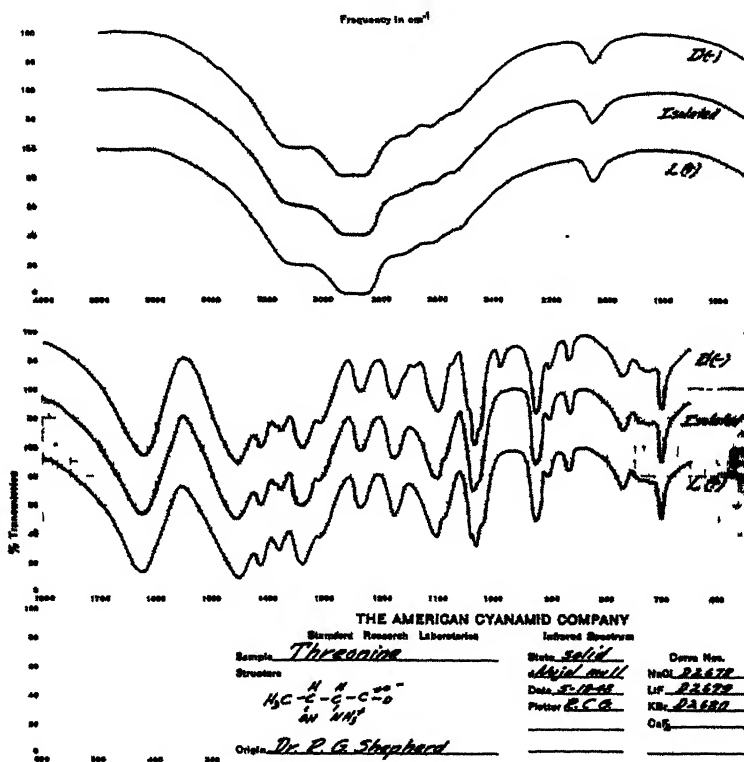


FIGURE 10 Spectra of *D* (-), isolated, and *L* (+) threonines

seen, both the isolated threonine and the *L* (+) threonine lack an absorption band near 990 cm^{-1} found in the *D* (-) isomer. From this, it appears as if the unknown is *L* (+) threonine.

These spectral differences observed in the case of *d*- and *l*-isomers were, at first thought, quite shocking. It is to be noted, however, that these materials were observed in the solid phase, with a spectrometer having some inherent polarization of its radiation. Because of the value of such observations, these studies are being pursued further.

Summary

- (1) The infrared spectra of eighteen antibiotic materials have been given.
- (2) The spectra of polymyxin and "Aerosporin" show differences in spite of their high molecular weights.
- (3) A fatty acid isolated from polymyxin appears to be similar to but not identical with pelargonic acid, showing evidence of more methyl groups than the latter.
- (4) *D* (+)- α,γ -diaminobutyric acid has been identified as a breakdown product of polymyxin.
- (5) *D*-serine has been identified as a hydrolysis product.
- (6) *L* (+) leucine has also been obtained.
- (7) *L* (+) threonine has been identified as a material isolated from polymyxin.

References

1. BARNES, R. BOWLING, R. C. GORE, E. F. WILLIAMS, S. G. LINSLEY, & E. M. PETERSEN. 1947. Infrared analysis of crystalline penicillins. *Anal. Chem.* **19**: 620-27.
2. BARNES, R. BOWLING, R. C. GORE, R. W. STAFFORD, & V. Z. WILLIAMS. 1948. Qualitative organic analysis and infrared spectrometry. *Anal. Chem.* **20**: 402-410.

PHARMACOLOGY OF POLYMYXIN*

By MORTON S. BRYER, EMANUEL B. SCHOENBACH AND ELEANOR A. BLISS

Department of Preventive Medicine, The Johns Hopkins University School of Medicine,
Baltimore, Maryland

Polymyxin (B-71) is an antibiotic derived from cultures of *Bacillus polymyxa*.¹⁻⁴ Its activity appears to be restricted to certain gram-negative species of bacteria. Aerosporin⁵⁻⁷ is a similar antibiotic derived from *B. aerosporus* Greer. It is the purpose of this presentation to describe some of the pharmacological studies which have been performed on experimental animals.

Toxicity. The acute toxicity of polymyxin was determined for albino Swiss mice by single subcutaneous injection and the L.D.₅₀ was 250 to 300 mg./kg. A one per cent solution of drug was used and occasional loss of fur and skin irritation at the injection site was noted with doses above 200 mg./kg. Deaths occurred within 6 hours with scratching of the head, ataxia, convulsions, paralysis, and respiratory arrest. Intraperitoneal injection of 250 to 275 mg./kg. produced convulsions followed by paralysis and death in 5 to 20 minutes. Stansly, Shepherd, and White² also recorded an L.D.₅₀ of 300 mg./kg. in mice receiving a single subcutaneous dose.

TABLE 1 reveals, in the outlined area, the doses of polymyxin in mg./kg.

TABLE 1
ACUTE TOXICITY OF "POLYMYXIN" IN MICE INJECTED WITH A SINGLE DOSE* (LOT NO. 7-7795)

No. of mice in group	Dose of "poly- myxin" (sub- cutaneously) mg./kg.	Deaths within 6 hours		Survivors 7 days	
		No.	Per cent total	No.	Per cent total
10	50	0	0	10	100
10	100	0	0	10	100
10	200	2	20	8	80
20	250	11	55	9	45
10	275	5	50	5	50
20	300	10	50	10	50
10	350	7	70	3	30

At toxic doses, symptoms noted were: ruffling of fur, ataxia, paralysis of hind legs, clonic convulsions, and respiratory depression.

* Albino swiss mice weighing 18-22 grams.

resulting in a mortality of 50 per cent in mice. Another lot tested had an L.D.₅₀ of 400-500 mg./kg.

As indicated in TABLE 2, dogs survived single, rapid intravenous injections of 10 and 15 mg./kg., administered over a period of 5 to 20 seconds. The 10 mg./kg. dose was associated with only transient dilatation of the

* These investigations were supported by grants received from Abbott Laboratories, Lederle Laboratories, Parke, Davis and Company, and the Upjohn Company.

TABLE 2
TOXICITY OF "POLYMYXIN" IN DOGS
Intravenous administration

Dosage mg./kg.	Dosage schedule	Total dose "polymyxin" mg.	Results observed
10	Rapid single injection	62.0	Dilated pupils, <i>survived</i> 8 days +
15	Rapid single injection	94.5	Dilated pupils, paralysis of hind legs, incontinence. <i>Survived</i> 8 days +
25	Rapid single injection	130.0	Convulsions, ascending paralysis, respiratory arrest. <i>Death</i> in 20 minutes
35	Slow 0.25 mg. per kilo per minute (pentobarbital anesthesia)	255.5	Respiratory depression, generalized flaccidity. <i>Death</i> at 150 minutes

pupils. Five seconds after receiving 15 mg./kg., a dog developed dilated pupils, scratching of the head, paresis of the hind limbs, and tremor. Incontinence and slowing of respiration followed. However, the animal recovered quickly and completely. Rapid injection of 25 mg./kg. was followed in 5 seconds by generalized convulsions, ascending paralysis, incontinence, and coma. In 7 minutes, apnea developed although heart pulsations remained of good quality. Death occurred in 20 minutes. Continuous intravenous drip administration at the rate of 0.25 mg./kg. per minute resulted in paralysis, apnea, and death when 35 mg./kg. had been given to a dog under light anesthesia.

Doses of 5 and 10 mg./kg. given intramuscularly twice daily for 7 days were tolerated by dogs with only occasional sneezing and scratching of the head. No weight loss was noted for one week following the last injection. Thus, large total amounts of drug produced no marked toxicity when administered over comparatively long periods of time.

Intrathecal injection of 1 mg. in the anesthetized dog was associated with no untoward reaction, while 5 mg. was followed by a transient paresis of the hind limbs which disappeared in one day. Administration of 10 mg. produced a similar muscular weakness which cleared completely over a 10-day period and was associated with a 2 plus Pandy reaction in the spinal fluid. In these experiments, the polymyxin was injected in 2 cc. of buffered saline, pH 7.4, diluted with 2 cc. of spinal fluid and given over a period of 5 to 10 minutes.

Young female rats observed to have no albuminuria detectable by precipitation with 12 per cent trichloroacetic acid, and no casts, were given 20 mg./kg. of polymyxin in single daily subcutaneous doses. Within 24 hours of the first injection, epithelial cells, cellular casts, and albumin were found in abundance in the urine. The specific gravity was noted to fall markedly but the urinary output was maintained and usually increased. The albuminuria and casts tended to decrease although the drug was continued. Simultaneous administration of a single subcutaneous dose of

TABLE 3
TOXICITY OF "POLYMYXIN" IN DOGS

Dosage mg. /kg.	Dosage schedule	Total dose "polymyxin" mg.	Results observed
Intramuscular administration			
5	Twice daily for 7 days	441.0	Scratching of nose and ears following occ. injection. No toxic reaction 7 days after last dose. Survived
10	Twice daily for 7 days	868.0	"
Intrathecal administration			
—	—	1.0	No apparent reaction
—	—	5.0	Transient paresis of hind legs
—	—	10.0	Paresis of hind legs. Improvement over 10-day period

crystalline *dl*-methionine 50 mg./kg. with the first injection of polymyxin and also as 1.0 per cent of the diet, or 1 gm./kg./day, did not prevent the urinary changes noted. No decrease in the hemoglobin or weight of the animals was observed.

Microscopic examination* of the kidneys of mice, dogs, and rats receiving polymyxin revealed damage to the tubular epithelium. Rats and dogs received 20 mg./kg. in a single daily subcutaneous dose for 4 days. Autopsy was performed on the fifth day. Microscopic sections of the kidneys showed occasional casts, and regeneration of tubular epithelium. Oral and subcutaneous doses of methionine, more than 50 times those of polymyxin, did not prevent these changes. Mice were given 250 to 350 mg./kg. in a single subcutaneous injection and sacrificed 48 hours later. The kidneys of these animals revealed marked necrosis of tubular epithelium and cast formation.

Levels in Body Fluids. Ninety minutes after single intramuscular doses of 5 and 10 mg./kg. in dogs, serum levels of 2.5 and 5.0 gamma/cc. were obtained. Three and a half hours after these injections, serum contained 2.5 and 1.25 gamma/cc., respectively. Levels obtained were approximately 4 times higher when dogs had been maintained on these doses twice daily for 7 days. These animals had considerable levels for a 5-hour period following the last injection. At the end of 23 hours, practically all the polymyxin had been eliminated from the serum.

Despite high blood serum levels obtained in dogs by both intravenous and intramuscular administration, no polymyxin was detected in the spinal fluid. This is noted at the top of TABLE 5, where blood serum levels greater than 320 gamma/cc. were associated with a spinal fluid in which no polymyxin was detected. Likewise, a dog received 5 mg./kg. twice daily for 2½ days, and although the blood serum contained 2.5 gamma/cc., no poly-

* Microscopic sections were prepared and their interpretation corroborated by Dr. Tobias Weinberg, Pathologist, Sinai Hospital, Baltimore, Md.

TABLE 4
ASSAY FOR "POLYMYXIN" IN BODY FLUIDS IN DOGS

Dosage schedule	Specimen examined	Interval post <i>last</i> injection	Level of "Polymyxin", gamma/cc.
Intramuscular administration			
5 mg./kg. single dose	Serum	90 minutes	2.5
		150 minutes	2.5
		210 minutes	2.5
10 mg./kg. single dose	Serum	90 minutes	5.0
		150 minutes	2.5
		210 minutes	1.25
5 mg. kg. b.i.d. for 7 days	Serum	15 minutes	10.0
		30 minutes	20.0
		60 minutes	20.0
		180 minutes	10.0
		5 hours	2.5
		23 hours	0.0
10 mg./kg. b.i.d. for 7 days	Serum	15 minutes	20.0
		30 minutes	20.0
		60 minutes	20.0
		180 minutes	20.0
		5 hours	2.5
		23 hours	0.125
5 mg./kg. b.i.d. for 5 injections	Serum	65 minutes	2.5
	Sp. fl.	65 minutes	0.0

TABLE 5
ASSAY FOR "POLYMYXIN" IN BODY FLUIDS IN DOGS

Dosage schedule	Specimen examined	Interval post <i>last</i> injection	Level of "Polymyxin", gamma/cc.
Intravenous Administration			
Continuous drip 0.25 mg./kg. per minute	Serum	150 minutes	320.0
	Sp. fl.	150 minutes	0.0
Intrathecal Administration			
1 mg. L ₄ -L ₅	Sp. fl.	45 minutes	10.0
		225 minutes	0.3
5 mg. cisternal	Serum	25 minutes	1.25
		200 minutes	0.625
	Sp. fl.	15 minutes	500.0
		295 minutes	20.0
10 mg. cisternal	Serum	20 minutes	0.0
		125 minutes	0.625
	Sp. fl.	15 minutes	320.0
		120 minutes	80.0

myxin could be determined in the spinal fluid. Brownlee *et al.*⁶ also reported that Aerosporin did not pass from blood to spinal fluid in normal animals.

Intrathecal administration of 1, 5, and 10 mg. resulted in spinal fluid levels of 10 to 500 gamma/cc. which fell to 0.3 to 20 gamma/cc. in 2 to 5 hours. Blood levels of 1.25 to 0.6 gamma cc. were obtained following intrathecal doses of 5 and 10 mgms.

Therapeutic Effectiveness. As indicated on the chart, a single subcutaneous injection of 1 mg./kg. of polymyxin protected the majority of mice infected intraperitoneally with 1000 lethal doses of *Klebsiella pneumoniae* type A* and *Hemophilus influenzae* type b. This dose was notably more effective than 1 mg./kg. of streptomycin in these experiments and pro-

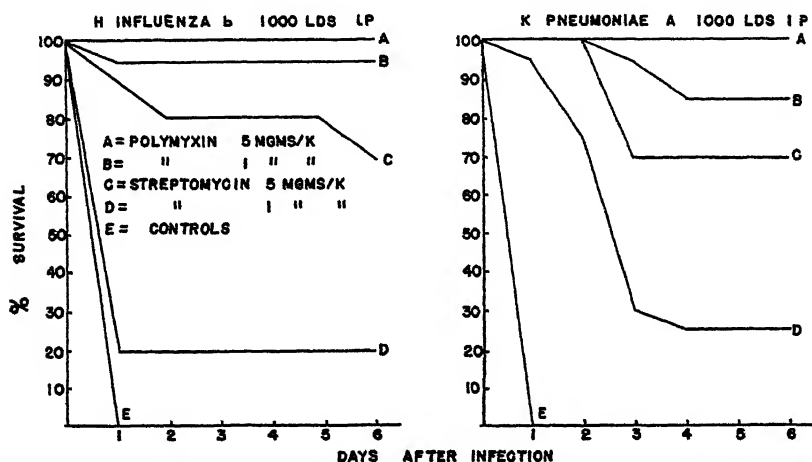


FIGURE 1. Comparison of polymyxin and streptomycin. 20 mice were used at each dosage level. Albino Swiss mice, 18-22 grams.

tected as well or better than 5 mg./kg. In another experiment, using twice the inoculum of *K. pneumoniae*, 1 mg./kg. of polymyxin gave as much protection as 10 mg./kg. of streptomycin. However, it is to be noted that 100 mg./kg. of polymyxin given subcutaneously in mice were unable to protect them from an intraperitoneal infection with gram-positive *Pneumococcus* type I.

Blood Cultures. Ten mice were infected with 1000 LD's of *K. pneumoniae* and cultures were taken from the tail blood at 2, 4, and 18 hours. Blood cultures, 2 hours post infection, were positive in all animals. At this time, 5 mice were given a single subcutaneous injection of 50 mg./kg. of polymyxin. All of the controls had positive cultures at 4 and were dead at 18 hours post infection. The 5 mice receiving the polymyxin had negative cultures at 4 and 18 hours, and 100 per cent survived for 6 days.

* *K. pneumoniae* type A, extremely virulent for mice, was obtained from Dr. H. J. White.

TABLE 6
EFFECT OF POLYMYXIN ON BLOOD CULTURES

Mouse	Drug	Dose mg./ kg.	Route of administration	Blood cultures						Per cent sur- vivors 6 days
				Hours after	Infection			Treatment		
					2	4	18			
					0	2	16			
1	Polymyxin	50	Subcutaneously	Mice infected intra- peritoneally with 1000 LD's of <i>K. pneumo- niae</i> type A and treatment de- layed 2 hours	+	-	-	100		
2	Polymyxin	50	Subcutaneously		+	-	-			
3	Polymyxin	50	Subcutaneously		+	-	-			
4	Polymyxin	50	Subcutaneously		+	-	-			
5	Polymyxin	50	Subcutaneously		+	-	-			
1	Control	—	—	Mice infected in- traperitoneally with 1000 LD's of <i>K. pneumo- niae</i> type A and untreated	+	+	Dead	0		
2	Control	—	—		+	+	Dead			
3	Control	—	—		+	+	Dead			
4	Control	—	—		+	+	Dead			
5	Control	—	—		+	+	Dead			

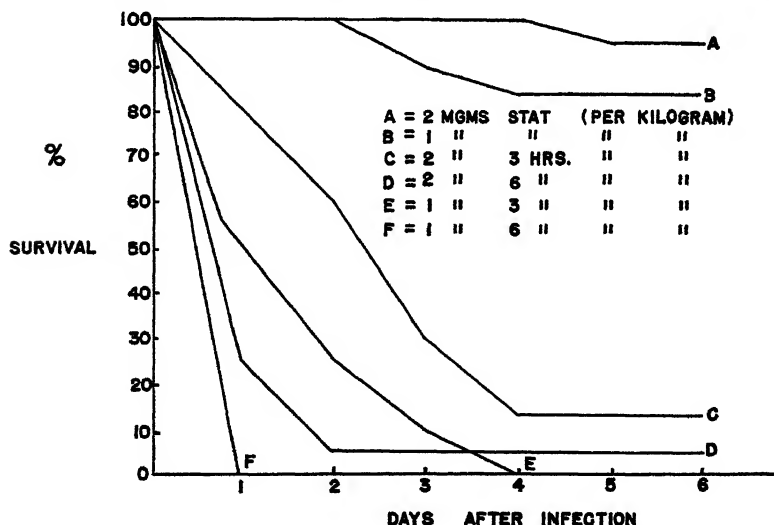


FIGURE 2. Time interval between infection and treatment. Mice infected with *K. pneumoniae* A, 1,000 LD's, treated with polymyxin subcutaneously. 20 mice were used at each dosage level. Albino Swiss mice, 18-22 grams.

Delay in Treatment. Delay in treatment of 3 or 6 hours following intraperitoneal infection with 1000 LD's of *K. pneumoniae* resulted in a reduction in the number of survivors as seen in the chart. Increasing the dose from 1 to 2 mg./kg. did not overcome the effect of delay in treatment.

Initial Dose. A single initial dose of polymyxin was more effective than the same amount given in divided doses. One mg./kg. of drug at the time of infection with 1000 LD's of *K. pneumoniae* protected 85 per cent of the mice. One-half mg./kg. immediately and 0.5 mg./kg. given in 3, 6, or 25 hours protected 70, 65, and 55 per cent, respectively. A quarter of a mg./kg. given immediately and in 3, 6, and 24 hours after infection resulted

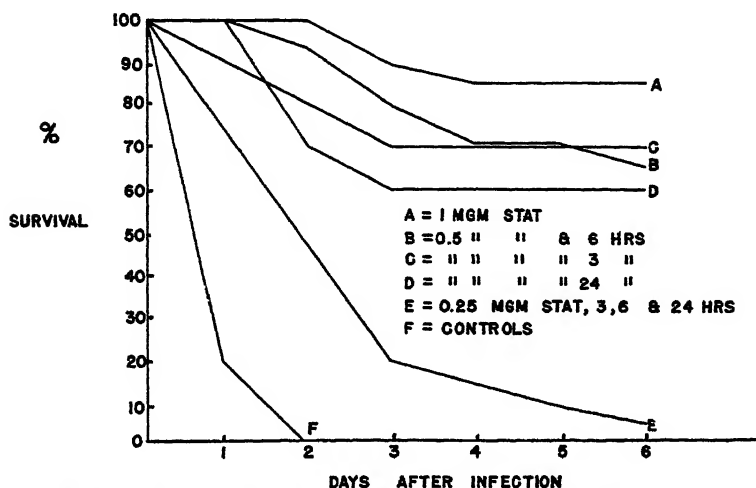


FIGURE 3. Variation of dose schedule. Mice infected with *K. pneumoniae* A, 1,000 LD's, treated with polymyxin subcutaneously. 20 mice were used at each dosage level. Albino Swiss mice, 18-22 grams.

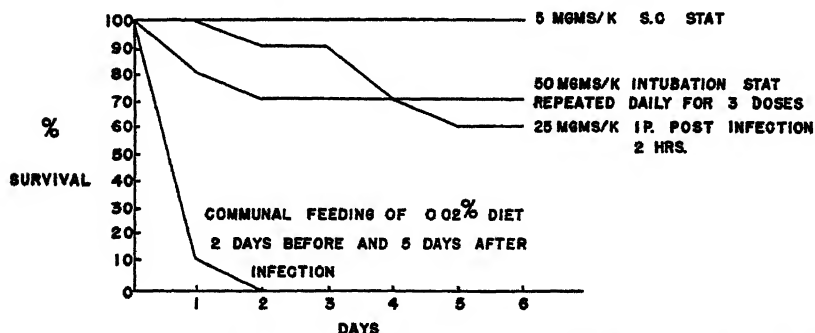


FIGURE 4. Routes of administration of polymyxin (infection of mice with *K. pneumoniae* A, 1,000 LD's) 20 mice were used at each dosage level. Albino Swiss mice, 18-22 grams

in only 5 per cent survivors. As suggested by *in vitro* experiments, the initial blood level may play an important role in therapy.

Routes of Administration. Different routes of administration of polymyxin were utilized in mice infected intraperitoneally with 1000 LD's of *K. pneumoniae*. Subcutaneous injection of 5 mg./kg. immediately following infection resulted in the highest percentage of survivors. Gastric

intubation of 50 mg. kg. immediately after infection and repeated daily for 3 doses resulted in 70 per cent survivors. Communal feeding of 20 mg. per cent of polymyxin mixed in the standard diet and fed 2 days before and 5 days after infection appeared ineffective. On the basis of the average daily intake of 5 grams per mouse, this diet should have supplied approximately 50 mg. kg., mouse, day.

Comparison of Polymyxin and Streptomycin. In this last figure, one may

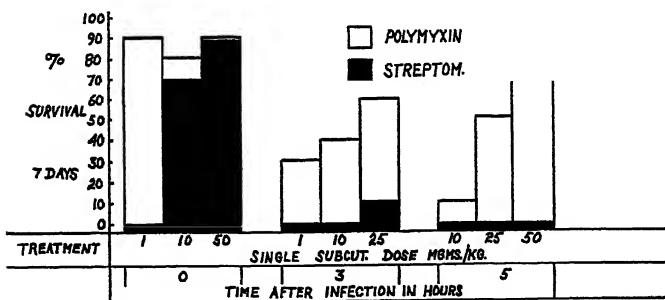


FIGURE 5. Comparison of delayed treatment of mice (intraperitoneal infection with *K. pneumoniae*, 10,000 LD's). 20 mice were used at each dosage level. Albino Swiss mice, 18-22 grams.

see represented graphically the comparative protection achieved with polymyxin and streptomycin given in a single subcutaneous injection immediately, 3 hours, and 5 hours after infection. Fifty mg./kg. of polymyxin injected 5 hours following intraperitoneal infection with 10,000 LD's of *K. pneumoniae* was needed to give as much protection as 1 mg./kg. administered immediately after infection. Fifty mg./kg. of streptomycin resulted in no survivors when given 5 hours post infection.

Summary

The acute toxicity of polymyxin was determined for mice by single subcutaneous injection, and the L.D.₅₀ was 250 to 300 mg./kg. for one lot and 400-500 mg./kg. for another.

Subcutaneous injections of 20 mg./kg. of polymyxin in young female rats was followed by albuminuria, cellular casts, and low specific gravity of urine. No decrease in weight or hemoglobin was noted. Microscopic sections revealed damage to the tubular epithelium of the kidneys.

The toxicity was greater when the drug was administered by rapid intravenous injection than when a slow intravenous injection was employed. Both methods appeared more toxic than the subcutaneous or intramuscular routes.

Adequate blood levels are readily obtained by subcutaneous and intramuscular injection. Polymyxin was slowly eliminated from the serum and tended to accumulate when 5 or 10 mg./kg. were administered twice daily.

Polymyxin does not seem to pass from the blood to the spinal fluid in normal animals.

Blood levels were obtained when polymyxin was administered intrathecally. Cisternal injection of 10 mg. in 0.25 per cent solution appeared to be irritating.

Subcutaneous administration was more effective in protecting mice than intubation. Communal feedings appeared to be ineffective.

The initial blood level may be an important factor in therapy.

While polymyxin is more toxic than streptomycin by weight, it is considerably more effective in the experiments performed.

Bibliography

1. BENEDICT, R. G. & A. F. LANGLYKKE. 1947. Antibiotic activity of *Bacillus polymyxa*. Soc. Amer. Bact. Abstr. of Proc. 24.
2. STANSLY, P. G., R. G. SHEPHERD, & H. J. WHITE. 1947. Polymyxin: A new chemotherapeutic agent. Bull. Johns Hopkins Hosp. 81: 43.
3. STANSLY, P. G. & M. E. SCHLOSSER. 1947. Studies on Polymyxin: Isolation and identification of *Bacillus Polymyxa* and differentiation of polymyxin from certain known antibiotics. J. Bact. 54: 549.
4. STANSLY, P. G. & M. E. SCHLOSSER. 1947. Studies on polymyxin: An agar diffusion method of assay. J. Bact. 54: 585.
5. AINSWORTH, G. C., A. M. BROWN, & G. BROWNLEE. 1947. "Aerosporin," an antibiotic produced by *Bacillus Aerosporus* (Greer). Nature 160: 263.
6. BROWNLEE, G., & S. R. M. BUSHBY. 1948. Chemotherapy and pharmacology of Aerosporin, a selective gram-negative antibiotic. Lancet I: 127.
7. SWIFT, P. N. 1948. Treatment of pertussis with Aerosporin. Lancet I: 133.

IN VITRO STUDIES OF POLYMYXIN*

By ELEANOR A. BLISS, CAROLINE A. CHANDLER, AND EMANUEL B. SCHOENBACH

Department of Preventive Medicine, The Johns Hopkins University School of Medicine, Baltimore, Maryland

Studies of the *in vitro* activity of polymyxin have been in progress at this laboratory since October, 1947. The preparation used is the hydrochloride of the partially purified concentrate, supplied by the Lederle Laboratories Division of the American Cyanamid Co., in 20-mg. ampoules labeled "B 71 Hydrochloride." Observations have been made on the stability of the agent to heat and changes in the hydrogen ion concentration, its range of action and potency, and its effect upon the growth of *Escherichia coli*. The likelihood of the development of drug fastness to polymyxin has been investigated. A method has been devised for assaying the concentration in body fluids and a search for antagonists has been started.

Environmental Factors. The medium used in most of the tests about to be described was Difco Heart Infusion Broth with 0.05 per cent dextrose added, but the constitution of the medium appears to be unimportant in titrations with polymyxin. The end-points are not raised by the presence of serum in the tests. In fact, with serum they have averaged somewhat lower in 17 instances when comparisons were made. However, heating polymyxin in the presence of serum does result in a raising of the end-points as much as 8 times, the average in 20 tests being 4.3 times. On the other hand, the agent, at pH 7.2 in broth or salt solution, appears to be quite stable to heat, there being little or no change in activity after it has stood at room temperature or in the 37°C. incubator for 18 hours, after heating at 56°C. for as much as 4 hours or after treatment in a boiling water bath for 10 minutes. Nevertheless, it keeps best, over long periods of time, in an acid environment in the deep freeze, -20°C.

The loss of activity upon heating in the presence of serum is not only puzzling, but also inconvenient. The sera of human beings and animals often contain antibodies for gram-negative bacteria which confuse the readings, and which must be eliminated by heating the sera at 56°C. for 30 minutes. In assaying for polymyxin in the blood, it is, therefore, essential to prepare a standard in serum and to heat it at the same time as the unknown sera. After heating, the standard and the unknown are diluted out in broth, in serial two-fold dilutions of $\frac{1}{2}$ cc. each, and the tubes are inoculated with $\frac{1}{2}$ cc. of a 1:10,000 dilution of a 20-hour old culture of *E. coli*. The titrations are read at 18 to 22 hours. The least amount of polymyxin which can be detected in this way is 0.62 to 1.25 $\mu\text{g./cc.}$

* These investigations were supported by grants received from Abbott Laboratories, Eli Lilly and Company, Lederle Laboratories Division, Parke, Davis and Company, and the Upjohn Company. The authors are indebted to Dr. Norman Weisman and Miss Jean Fisher for the Lipositol, which they prepared, and to Mrs. Elizabeth Burr for technical assistance in the first part of the study.

In broth, without serum, the end-points are quite reproducible. Ninety per cent of the times in our tests the last clear tube has been either the one containing 0.31 $\mu\text{g./cc.}$ or that containing 0.156 $\mu\text{g./cc.}$ We have however, encountered "skipping," illustrated in TABLE 1, with considerable frequency.

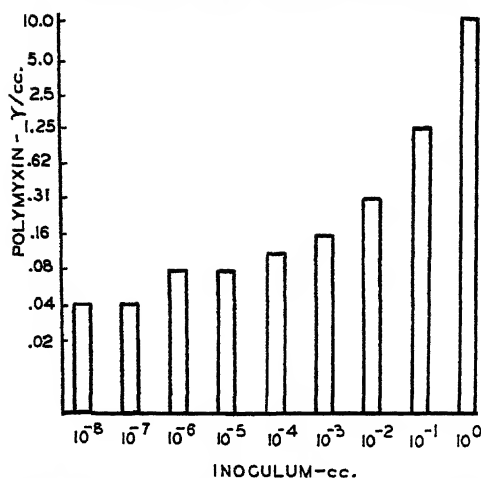
TABLE 1
"SKIPPING"—AN IRREGULARITY OBSERVED IN THE COURSE OF TITRATIONS OF POLYMYXIN BY THE BROTH DILUTION METHOD
Test organism: *E. coli*, 24-hour reading

	Concentration of polymyxin mg./cc.							
	5.0	2.5	1.25	0.62	0.31	0.156	0.08	0
Usual result	—	—	—	—	—	+	+	+
Occasional result	—	—	—	+	—	+	+	+
	0.7	0.6	0.5	0.4	0.3	0.2	0.1	0
Usual result	—	—	—	—	—	+	+	+
Occasional result	—	+	+	—	—	+	+	+

Since the skip always occurs above the usual end-point, it is evident that it is not caused by the presence of an enhancing substance in the clear tube. The possibility was entertained that the growth in the "false positive" tube might be the result of resistant organisms growing out. However, titrations with these organisms showed that they were as sensitive to polymyxin as the parent culture. Moreover, all attempts to encourage the development of resistance to polymyxin in normally sensitive strains have ended in the same failures in this laboratory as have been noted by other investigators.¹ Polymyxin has not given rise to resistant variants.

Since the phenomenon of "skipping" was not caused by the outgrowth of resistant organisms, nor by the presence of additional antibacterial substances in the clear tube, it seemed that it must be due to the presence, in the "false positive" tube, of material which interfered with the activity of polymyxin. Investigation of possible contaminants in the test tubes finally showed that soap was the probable culprit. The anti-polymyxin effect of soap will be described in greater detail. It is enough to say here that, when special pains were taken with the rinsing of the laboratory glassware, the frequency of "skipping" was greatly reduced.

Although titrations with polymyxin are not affected by differences in the medium, they are extremely sensitive to differences in the size of the inoculum. As shown in FIGURE 1, the amount of polymyxin required to suppress the growth of *E. coli* increases two-fold for every hundred-fold increase in the inoculum for inocula below 0.1 cc. of culture (per 10 cc. of broth), but an inoculum of 0.1 cc. requires for inhibition four times as much drug as an inoculum of 0.01 cc., and an inoculum of 1 cc. requires 8 times as much as 0.1 cc. A standard inoculum of approximately 200,000 organisms per cc. has therefore been adopted in this laboratory. It is achieved by the use of 0.1 cc. of a 1:1000 dilution or 0.5 cc. of a 1:10,000 dilution in tests in which the final volume is 1 cc.

FIGURE 1. Minimal inhibitory concentrations of polymyxin for increasing inocula of *E. coli*.

Range of Action. The observation of others³ that polymyxin is active in low concentration against a number of gram-negative bacteria has been confirmed. In TABLE 2 are shown the minimal inhibitory concentrations

TABLE 2
MINIMAL INHIBITORY CONCENTRATIONS OF POLYMYXIN FOR VARIOUS GRAM-NEGATIVE BACTERIA

Organism	No. of strains	Polymyxin μg./cc.
<i>Aerobacter</i>	5	0.3
<i>Aerobacter</i>	1	0.6
<i>Aerobacter</i>	1	1.2
<i>E. coli</i>	2	0.3
<i>E. communior</i> ?	1	2.5
"Paracolon"	1	0.16
<i>E. typhosa</i>	1	0.08
<i>K. pneumoniae</i>	3	0.3
<i>Ps. aeruginosa</i>	3	2.5
<i>H. influenzae</i>	1	0.3
<i>Proteus vulgaris</i>	1	50.0
<i>Proteus vulgaris</i>	2	more than 100
<i>Meningococcus</i>	1	more than 30

for 1 to 7 strains of several genera. The only resistant bacteria encountered were the 3 strains of *Proteus vulgaris* and the *Meningococcus*. The

test with the last organism, however, is not comparable to the others because it was made on a solid medium while the others were run in broth.

Mechanism of Action. Titrations with polymyxin are remarkable for the stability of the end-points. Prolonging the incubation time of the tests rarely results in an upward shift of the reading, and sub-cultures from the tube containing the minimal inhibitory concentration of the drug usually show no viable organisms. These observations suggested that polymyxin is bactericidal in its action and not merely bacteriostatic. To determine this point, studies were made of the effect of the agent upon the rate of multiplication of *E. coli*. Ten-cc. lots of broth containing various amounts of polymyxin were inoculated with 0.1 cc. of a 1:100 dilution of a 20-hour-old culture. After thorough mixing, plates were poured immediately and after 1, 3, 6, and 24 hours of incubation, and the colonies were counted.

The results of one such test are shown in FIGURE 2. As little as 0.05

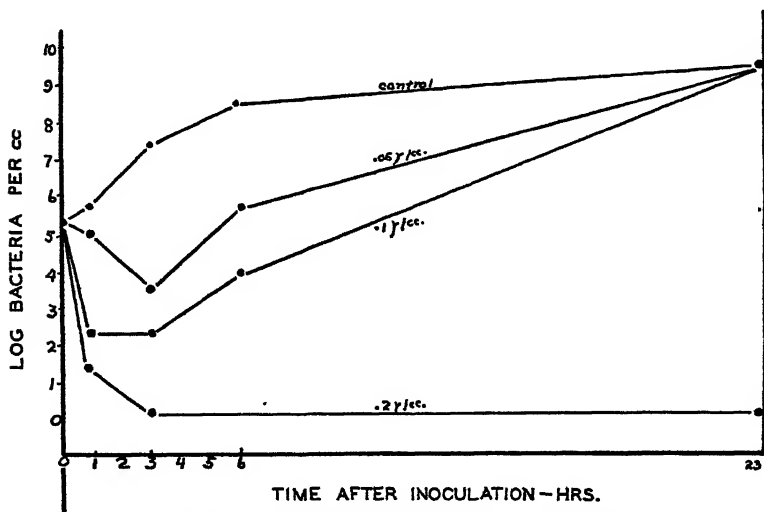


FIGURE 2. Effect of polymyxin on the multiplication of *E. coli*.

µg./cc. polymyxin had a definite effect upon the rate of multiplication of *E. coli*; at 3 hours there were only one-hundredth as many organisms present as at the start. The effect of 0.1 µg./cc. was more pronounced, 99.9 per cent of the bacteria having disappeared in the first hour. As with the smaller amount of polymyxin, however, the remaining viable organisms, about 200 per cc., recovered and, within 24 hours, growth was as heavy in this culture as in the drug-free control. In the presence of 0.2 µg./cc. the 20 organisms per cc. which were still alive at 1 hour had disappeared in 3 hours, and plates poured at 24 hours with 0.1 cc. continued to show no growth.

The speed with which polymyxin acts is striking. When larger amounts of the agent are used, one often finds that half the bacterial population has

disappeared before one has made the initial plate. Such rapid action suggests that it is not necessary for the bacteria to be dividing in order to be susceptible to the drug. However, comparative growth studies carried out at 37°C. and at 10°C showed (FIGURE 3) that, while polymyxin has a

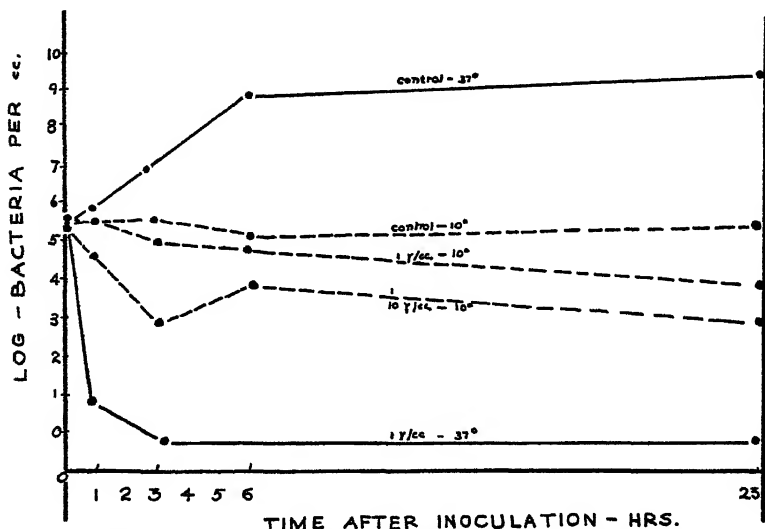


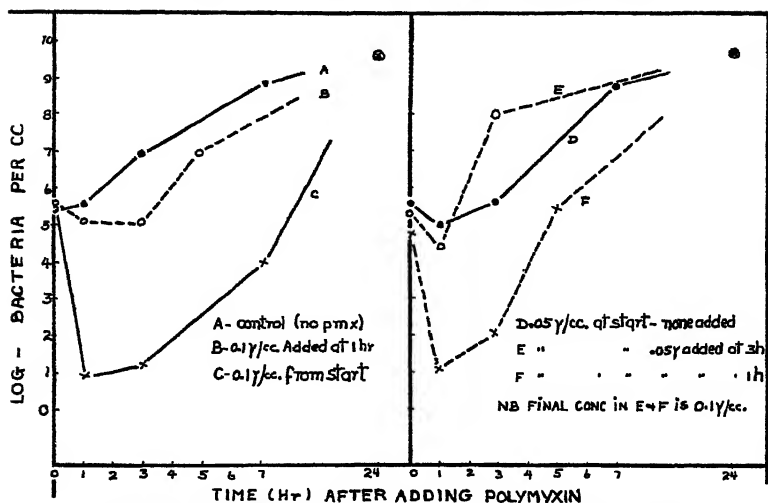
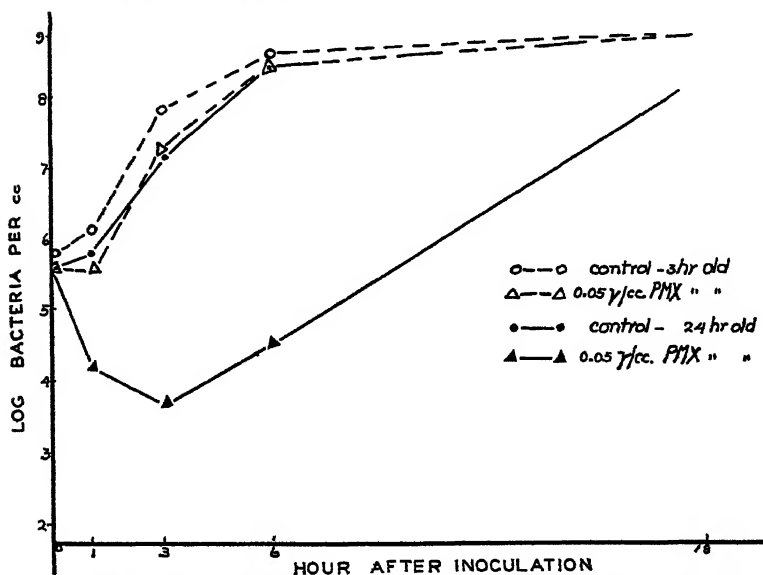
FIGURE 3. Effect of polymyxin on the multiplication of *E. coli* at 37°C. and 10°C.

slight effect in the cold, it is very much less than that observed in the usual tests at 37°C., and also indicated that, like the other antibiotic agents, polymyxin acts only on the dividing cell.

On the other hand, that the organisms must not be in too active a phase of growth was shown by two other series of observations, illustrated by FIGURES 4 and 5. In the first place, it was found that the addition of polymyxin at intervals during a growth study had less and less effect the later the addition was made, even when precautions were taken to insure the presence of approximately equal numbers of bacteria at the various times of addition. Then a direct comparison of the effect of the agent on the multiplication of equivalent inocula of 3- and 24-hour-old cultures showed it to be almost devoid of activity for the young culture.

The recovery of cultures which, like the old culture containing polymyxin shown in FIGURE 5, seem at 3 or 6 hours to be on the way to extinction is interesting. It cannot be explained by the outgrowth of resistant variants, since resistance to polymyxin does not occur, although it might represent a temporary adaptation to the drug. The question is still unanswered, but it led to a search for antagonists to polymyxin, first among bacteria and their products and later elsewhere.

First, it was found that the growth of *E. coli* in polymyxin, as occurs when large inocula are employed, resulted in a loss of activity on the part of the agent. A similar loss was sustained when solutions were treated for 1

FIGURE 4. Effect of late addition of polymyxin on multiplication of *E. coli*.FIGURE 5 Effect of polymyxin on multiplication of old and young cultures of *E. coli*.

hour at 37°C. with washed, resting, saline suspensions of the same organism, suggesting that the diminished activity of the drug was the result of its adsorption on the bacterial bodies rather than evidence of neutralization by a product of bacterial growth. On the other hand, treatment of solutions with a living non-susceptible organism, the C203 strain of hemolytic streptococcus, caused no loss of activity whatever.

Peptone, para-aminobenzoic acid, and methionine were tested for an-

tagonism to polymyxin. The first two had no effect, and methionine, the only amino acid so tested to date, raised the end-point only one tube, an alteration which is within the limits of experimental error.

As mentioned earlier, it was suspected that the presence of soap in the test tubes might account for the "skipping" observed in titrations of polymyxin. This hypothesis was investigated. A 500 mg./cc. solution of the soft green soap used in this laboratory was prepared in water and sterilized by boiling. Serial ten-fold dilutions in water were made and 0.1 cc. of each dilution was pipetted into a row of Wassermann tubes. To each was added 0.5 cc. of one of a series of two-fold broth dilutions of polymyxin. All of the tubes were inoculated with 0.4 cc. of a 1:8000 dilution of *E. coli*. As shown in TABLE 3, concentrations of soap from 5 to 5,000 $\mu\text{g./cc.}$ inter-

TABLE 3
NEUTRALIZATION OF POLYMYXIN BY SOAP, ASO-LECTIN AND LIPOSITOL
Test organism: *E. coli*, 200,000/cc.

"Antagonist"	Concentration of antagonist— $\mu\text{g./cc.}$						
	0.05	0.5	5.0	50	500	5000	50,000
	Amount of polymyxin neutralized— $\mu\text{g.}$						
Soap	0	0	.16	1.0	20	320	0*
Aso-lectin	0	0	0	.16	5	20	—
Lipositol	0	0	.16	1.0	10	80	—

* This concentration of soap inhibited *E. coli*.

fered with increasing amounts of polymyxin up to 320 $\mu\text{g./cc.}$ Soap, at 50,000 $\mu\text{g./cc.}$, was itself antibacterial.

Next, aware that polymyxin was a basic polypeptide⁸ and that such substances are interfered with by phosphatids of the lecithin type, a soybean lecithin preparation, Aso-lectin, was tested in much the same manner as soap. Finally, tests were made with lipositol, a phospholipid component of soybean lecithin. Both of these substances interfered with polymyxin. Rather disappointingly, the lipositol was little more effective in this respect than the parent substance. However, lipositol is a somewhat unstable compound and, although it was used within 24 hours of the time when the solutions were prepared, it is possible that it may have lost in activity through oxidation.

In being interfered with by soap, Aso-lectin, and lipositol, polymyxin behaves like the cationic detergents. This observation, however, does not explain the antibacterial activity of polymyxin. The cationic detergents are far more effective against gram-positive than against gram-negative bacteria, while polymyxin is active only against the latter.

Summary

Polymyxin is an antibiotic agent, effective, in low concentration against many gram-negative bacteria. Its activity is not greatly influenced by

environmental factors, although heating it in the presence of serum proved to be deleterious. It is sensitive to changes in the size of the inoculum. Three strains of *B. proteus* and a single strain of *Meningococcus* appeared to be naturally resistant to the antibiotic, but repeated exposure to it does not result in the development of resistance in normally sensitive strains.

Polymyxin is rapidly bactericidal, but it is more effective at 37°C. than at 10°C., and against old rather than young cultures. Its action is impaired by substances which antagonize the kationic detergents.

References

1. STANSLY, P. G., R. G. SHEPHERD, & H. J. WHITL. 1947. Bull. Johns Hopkins Hosp. 81: 43-54.
2. STANSLY, P. G. & M. E. SCHLOSSER. 1947. J. Bact. 54: 549-556.
3. SCBBAROW, Y. Personal communication

THE PHARMACOLOGY OF POLYMYXIN A, B, AND D

By GEORGE BROWNLEE, S. R. M. BUSHBY, AND E. I. SHORT.

The Wellcome Physiological Research Laboratories, Beckenham, Kent, England

A pharmacological study of a classical kind, as applied to impure materials of biological origin, is a hazardous and probably profitless exercise, and it is not with this aspect that we propose to deal. Thus, many interesting side effects, often varying from one preparation to another and probably associated with impurities, will receive a bare mention.

Polymyxin A is the antibiotic described by Ainsworth, Brown, and Brownlee¹ and Brownlee and Bushby² and previously known as "Aerosporin." It is produced by *Bacillus polymyxa* strain C.N.1984 and contains the amino acids *D*-leucine, *L*-threonine, and *L*- α , γ -diaminobutyric acid, together with an optically active fatty acid of known chemical and physical characteristics, $C_{18}H_{35}O_2$, but, as yet, incompletely characterised.

Polymyxin B is the antibiotic produced by *B. polymyxa* strain C.N.1419 and contains leucine, phenylalanine, threonine, and diaminobutyric acid, together with the same fatty acid found in polymyxin A—but perhaps in different amount.

Both differ chemically from polymyxin D ("Polymyxin," Stansly *et al.*³) in that the latter contains the amino acid, serine, in addition to leucine, threonine, diaminobutyric acid, and the identical fatty acid. Polymyxin D contains no phenylalanine.

The composition of a fourth member, polymyxin C, has been described. It differs from the other three in having no leucine but contains phenylalanine, threonine, diaminobutyric acid, and the optically active $C_{18}H_{35}O_2$ fatty acid.

The amino acid composition of the four antibiotics is shown in TABLE 1.

TABLE 1
THE AMINO ACID COMPOSITION OF POLYMYXIN A, B, C AND D DERIVED FROM STRAINS OF *B. polymyxa*

	A	B	C	D
Leucine	+	+	—	+
Phenylalanine	—	+	+	—
Threonine	+	+	+	+
Serine	—	—	—	+
α , γ -Diaminobutyric acid	+	+	+	+

Throughout the present report, the concentrations of polymyxin A are stated in terms of "Polymyxin A hydrochloride Standard 1947," containing 10,000 "units" per mg. Material of this standard may, for all practical purposes, be considered to be the pure hydrochloride. Polymyxin B has not yet been obtained as a pure substance and its intrinsic potency is unknown. Preliminary studies indicate, however, that its potency will not

differ materially from that of polymyxin A. Polymyxin B has, therefore, been expressed as "units" of polymyxin A. The assay procedure is the short turbidimetric method using *E. coli* (Brownlee and Bushby²). Polymyxin D is referred to in terms of pure antibiotic of 2,000 units per mg. (Stansly *et al.*³).

IN VITRO Antibacterial Activity

The minimal concentrations of polymyxin A, B, and D which inhibited growth of large inocula of a range of organisms were determined by streaking 24-hour old broth cultures of the organisms on nutrient agar incorporating 2 per cent horse blood and containing two-fold dilutions of the antibiotics. Repetition, using the same strains, show slight variation in the minimum inhibitory concentration. A typical comparison is given in

TABLE 2
COMPARATIVE MINIMAL CONCENTRATIONS OF POLYMYXIN A, B, AND D INHIBITING LARGE INOCULA

Culture	Organism	Polymyxin		
		A	B	D
C.N.		µg./ml.	µg./ml.	µg./ml.
512	<i>Eberthella typhosa</i>	0.25	0.25	0.30
312	<i>Escherichia coli</i>	0.25	0.25	0.30
740	<i>Brucella bronchiseptica</i>	0.125	0.125	0.15
247	<i>Vibrio comma</i>	0.25	0.25	0.30
164	<i>Salmonella schottmuelleri</i>	0.125	0.25	0.30
1520	<i>Shigella paradysenteriae</i>	0.016	0.03	0.04
214	<i>Brucella abortus</i>	1.0	1.0	1.25
144	<i>Aerobacter aerogenes</i>	0.25	0.25	0.45
1833	<i>Pasteurella muriseptica</i>	0.25	0.25	0.45
200	<i>Pseudomonas aeruginosa</i>	1.0	2.0	1.5
181	<i>Klebsiella pneumoniae</i>	2.0	2.0	2.5
13	<i>Streptococcus viridans</i>	>4.0	>4.0	>4.0
491	<i>Staphylococcus aureus</i>	>4.0	>4.0	>4.0
10	<i>Streptococcus pyogenes</i>	>4.0	>4.0	>4.0
329	<i>Proteus vulgaris</i>	>4.0	>4.0	>4.0
127	<i>Haemophilus pertussis</i> *	0.25	0.25	0.15

Polymyxin A 10,000 units per mg. (BROWNLEE & BUSHBY²); B assayed in terms of A; D 2,000 units per mg. (STANSLY *et al.*³).

* Bordet-Gengou.

TABLE 2. No significant difference in the sensitivity of these organisms to the three antibiotics is detected by the method using large inocula.

Effect of the Size of the Inoculum

Polymyxin B is bactericidal like polymyxin A and D. Time, number of organisms, and concentration appear to bear a simple relation one to the other. The fact that the antibacterial action of the three antibiotics depends on the number of organisms present is illustrated in TABLE 3. This shows the effect of two-fold dilutions of the antibiotics in nutrient broth, inoculated with 0.05 ml. per 5 ml. of a 24-hour broth culture of *E. typhosa* diluted to 10^0 , 10^{-2} , 10^{-4} , and 10^{-6} . The cultures were incubated at 37°C.

TABLE 3
THE ANTIBACTERIAL EFFICIENCY OF POLYMYXIN A, B AND D, VARYING WITH THE NUMBER
OF ORGANISMS (THIS IS LESS SO WITH D)

Poly- myxin	Inoc.	2 0	1.0	0 5	0.25	0.125	0.06	0 03
A	10^0	—	+	+	+	+	+	+
	10^{-1}	—	—	—	—	+	+	+
	10^{-2}	—	—	—	—	—	+	+
	10^{-3}	—	—	—	—	—	—	+
B	10^0	—	+	+	+	+	+	+
	10^{-1}	—	—	—	—	+	+	+
	10^{-2}	—	—	—	—	—	+	+
	10^{-3}	—	—	—	—	—	—	+
D	10^0	+	+	+	+	+	+	+
	10^{-1}	—	—	+	+	+	+	+
	10^{-2}	—	—	—	+	+	+	+
	10^{-3}	—	—	—	—	+	+	+

for 18 hours. Polymyxin A and B are similarly influenced by the number of bacteria present; D is less so.

IN VIVO Antibacterial Activity

The chemotherapeutic value of Polymyxin A, B, and D was assessed in a side-by-side comparison, in groups of mice infected with Gram-negative pathogens.

Haemophilus pertussis. Groups of mice, infected with 0.05 ml. of a broth suspension containing 10^7 organisms (Kendrick *et al.*⁴) washed from a 48-hour culture on Bordet-Gengou medium and constituting about 10,000 average lethal doses, die consistently within four to seven days. With such groups, the relative chemotherapeutic effects of polymyxin A, B, and D were compared for a period of 14 days, during which survivors were recorded every second day. A protocol of a typical experiment is reproduced in TABLE 4, in which the average survival time is calculated by adding together the number of days survived by each mouse and dividing the sum by the number of animals in the group. This table shows polymyxin A and B to be similar and to be more effective than polymyxin D.

Eberthella typhosa. Mice infected intraperitoneally with 0.5 ml. of a suspension of Rawlings strain, containing 1000 average lethal doses in 5 per cent mucin, develop an overwhelming infection, so that most of them die within six hours. A typical protocol of a comparison among the three antibiotics in these test animals is reproduced in TABLE 5. The antibiotics were administered immediately after infection, six hours later, and twice daily for the following three days with subcutaneous injections of varying doses of the antibiotics.

Polymyxin B and D show similar protection, and appear to be less efficient than polymyxin A.

TABLE 4

PROTECTIVE ACTION OF POLYMYXIN A, B, AND D TO MICE INFECTED WITH 10,000 LETHAL DOSES OF *H. pertussis*

Polymyxin	Dose	No. doses	No. mice	No. mice surviving on day				Average survival rate: days	Per cent survival
				2	6	8	14		
A	mg.								
	0.1	2	36	36	36	35	28	13.1	77.8
	0.05	6	66	66	66	65	60	13.6	91.0
	0.025	6	36	36	32	28	20	10.8	55.5
B	0.1	6	6	6	6	6	6	14.0	100.0
	0.05	6	11	11	11	10	9	12.9	81.8
	0.025	6	13	13	12	8	6	10.2	46.0
D	0.25	6	23	23	23	23	21	13.7	91.3
		2	21	21	21	19	12	12.3	57.0
	0.125	6	19	19	19	16	11	11.8	57.9
	0.1	2	14	14	14	11	3	9.6	21.4
	0.625	6	21	21	20	17	9	10.6	43.0
Lethal controls	—	—	48	48	21	3	0	4.9	0.0

Polymyxin A as "pure" polymyxin A 10,000 units/mg. (BROWNLEE & BUSHEY²); Polymyxin B in terms of A; D as "pure" polymyxin D 2,000 units/mg. (STANLEY *et al.*³).

TABLE 5

THE EFFECT OF POLYMYXIN A, B AND D IN PROTECTING MICE INFECTED WITH 1,000 LETHAL DOSES OF *E. typhosa*

Polymyxin	Dose	No. of orgns.	No. of mice	No. mice surviving on day				Average survival rate: days	Per cent survival
				1	3	5	7		
A	mg.							days	
	0.1	5×10^7	18	18	18	18	18	7.0	100
	0.05	5×10^7	18	18	17	16	14	6.4	76.7
B	0.1	5×10^7	18	17	13	13	13	5.3	72.2
	0.05	5×10^7	18	16	16	13	12	5.5	66.7
D	0.2	5×10^7	12	12	12	11	11	6.9	91.7
	0.1	5×10^7	12	12	9	9	9	5.6	75.0
Lethal control		5×10^7	18	1	—	—	—	0.0	0.0
		5×10^8	6	4	—	—	—	1.0	0.0
		5×10^4	6	3	3	3	3	3.5	50.0

Polymyxin A is "pure", 10,000 units per mg. (BROWNLEE & BUSHEY²); B is in terms of A; D is "pure" 2,000 units per mg. (STANLEY *et al.*³).

Klebsiella pneumoniae. Protection experiments were designed also with groups of mice infected intraperitoneally with 10,000 and 1,000 lethal doses of *K. pneumoniae* suspended in 0.5 ml. of 5 per cent mucin. They were treated immediately, 6 hours later, and twice daily for the following 3 days with subcutaneous injections of varying doses of polymyxin A, B and D. Protection against this strain of the organism is not high (TABLE 6), and it is

TABLE 6

THE EFFECT OF POLYMYXIN A, B AND D IN PROTECTING MICE INFECTED WITH 10,000 AND 1,000 LETHAL DOSES OF *K. pneumoniae*

Polymyxin	Dose	No. of orgns	No. of mice	No. mice surviving on day				Average survival rate: days	Per cent survival
				1	3	5	7		
A	0.1	5×10^8	6	5	1	1	1	1.8	16.7
		5×10^7	6	5	2	2	1	3.8	16.7
	0.05	5×10^8	6	6	3	3	3	4.0	50.0
		5×10^7	6	6	3	3	3	4.0	50.0
B	0.1	5×10^8	6	1	—	—	—	0.0	0.0
		5×10^7	6	6	5	4	4	4.8	66.7
	0.05	5×10^8	6	2	—	—	—	0.0	0.0
		5×10^7	6	6	5	2	2	3.3	33.3
D	0.2	5×10^8	6	6	2	1	1	2.5	16.7
		5×10^7	6	6	2	2	2	3.3	33.3
	0.1	5×10^8	6	6	1	1	1	2.5	16.7
		5×10^7	6	6	3	3	2	2.5	33.3
Lethal controls		5×10^8	6	0	—	—	—	0.0	0.0
		5×10^7	6	0	—	—	—	0.0	0.0
		5×10^6	6	6	2	2	2	3.0	33.0
		5×10^4	6	6	4	4	4	5.0	66.7

Polymyxin A, in terms of "pure" substance, 10,000 units per mg. (BROWNLEE & BUSHBY²); B in terms of A; D, in terms of "pure" substance, 2,000 units per mg. (STANLEY *et al.*³).

evident that the higher infecting dose was overwhelming. Comparison at the lower infecting dose shows the relative efficiency of the three antibiotics to be close.

Toxicity

Acute Toxicity to Mice. The acute toxicity to mice was studied after the intravenous injection of varying doses of polymyxin A, B and D into groups of 10 animals, weighing 18–20 g. With lethal doses of polymyxin A and B, death occurs from respiratory failure in less than two minutes. With near-lethal doses there is vasoconstriction, muscular incoordination, and respiratory distress. This stage is followed by clonic convulsions and then flaccid paralysis associated with respiratory embarrassment and cyanosis, but the animals recover within 10 minutes.

The symptoms seen with polymyxin D are similar. Differences are that recovery may be delayed for 30 minutes or more, the paralysis is much less evident, and deaths may be delayed sometimes as long as 30 minutes. The acute intravenous toxicities of several batches of polymyxin A and B and of three batches of polymyxin D have been determined. Inspection of the slopes relating probits corresponding to mortalities and the logarithms of the doses used for all three antibiotics show them to be of the same order. There is no evidence of more than one agent contributing to the death of the mice.

The mean intravenous L.D.₅₀ for polymyxin A corresponds to 6.9 mg./kg. of "Polymyxin A Standard (1947)," with a range of 4.2 to 13.5, mean of estimates on 21 different samples.

The mean intravenous L.D.₅₀ for polymyxin B corresponds to 6.1 mg./kg. of "Polymyxin A Standard (1947)," range 4.2 to 10.7, mean of estimates on 12 different samples.

The mean intravenous L.D.₅₀ for polymyxin D corresponds to 11.9 mg./kg., mean of estimates on 3 samples, with a range of 8.7 to 18.0.

Acute Toxicity to Mice—Intraperitoneal Route. The acute toxicity after injection by the intraperitoneal route was observed in groups of mice, the antibiotics being dissolved in 0.4 ml. quantities of distilled water. The toxic signs were essentially similar for all antibiotics and were evident 30 minutes after injection, with weakness of hind limbs, followed by shivering and loss of temperature control. Clonic convulsions developed in 15–30 minutes, with respiratory embarrassment, and death from respiratory failure occurred after 2–4 hours. Recovery from near toxic doses occurs in 3 to 4 hours.

The mean intraperitoneal L.D.₅₀ for polymyxin A corresponds to 13.9 mg./kg. "Polymyxin A Standard (1947)" with a range of 10.5 to 19.0 for a mean of five samples.

The mean intraperitoneal L.D.₅₀ for Polymyxin B corresponds to 12.1 mg./kg. "Polymyxin A Standard (1947)" with a range of 8.7 to 17.5, mean of 11 samples.

The mean intraperitoneal L.D.₅₀ for polymyxin D corresponds to 24.5 mg./kg., a mean of two samples, range 24.0–25.0.

Acute Toxicity to Mice—Subcutaneous Route. The acute toxicity after subcutaneous injection of the three antibiotics was observed in groups of mice. The toxic signs were similar to those observed after administration by other routes, but were delayed for 30 minutes to 1 hour. Deaths occurred until 36 hours and the estimates were made at 48 hours. For polymyxin A, the mean L.D.₅₀ corresponded to 87.5 mg. per kg. with a range of 79.2 to 87.5 mg. per kg. a mean of 3 samples.

For polymyxin B, the mean L.D.₅₀ was 82.5 mg. per kg. with a range of 70.5 to 92.4 mg. per kg. for one sample.

For polymyxin D, the mean L.D.₅₀ was 160 mg. per kg. with a range of 112.5 to 220.0 mg. for two samples.

The many-fold differences between the acute toxic doses for these antibiotics by the intravenous or intraperitoneal routes, when compared with their toxicities by the subcutaneous route, suggests a local tissue fixation.

Acute Toxicity—Intracisternal—Rabbit. Groups of rabbits injected intracisternally with polymyxin A contained in 0.25 ml. isotonic buffer show respiratory depression and respiratory failure; other central effects appear to be absent.

The average lethal dose appears to be 0.6 mg./kg. with a range of 0.5 to 0.75, a mean of 8 observations.

Vestibular Function. The injection intracisternally in rabbits of sublethal doses of polymyxin A, in the range of 0.3 to 0.5 mg./kg., reveals no evidence of vestibular dysfunction. A temporary, infrequent, rotatory nystagmus, such as follows the injection of sterile saline, is seen during the first hour. Thereafter, nystagmus after rotation in the lateral plane is normal during periods of 12 days observation. Animals similarly treated with 3 mg./kg. of pure Streptomycin sulfate show, for periods of about 4 hours, severe respiratory depression, extensor tremors, stupor, and inability to hold the head horizontal, although jerking attempts were made to do so. However, other signs of vestibular dysfunction, such as rate and duration of nystagmus, were unaltered.

Circulatory Effects. In cats anesthetized with pentobarbitone sodium, traces of depressor substances were detected in impure batch preparations of polymyxin A and B. The amount of depressor substance, which is readily removed by further purification, is of the order of 0.2 to 3 μ g. equivalent of Histamine per mg. of 50 per cent pure antibiotic.

Kidney-Damaging Factor. Polymyxin A preparations were shown to contain a nephrotoxic principle (Brownlee and Bushby⁷). Since nephrotoxic action and purity were not directly related from batch to batch, it appeared possible that the principle was not necessarily also the antibiotic. Indeed, at the present time, the position remains unproved, yet the purest material available to us produces substantial proteinuria in the experimental animal and in man. Meanwhile, the observation that the closely allied polymyxin B is free from the nephrotoxic principle, although itself impure, offers an alternative explanation for the batch variability. The lesions produced by polymyxin A preparations, when studied histologically in the rat and guinea pig, appear to be restricted to the distal portion of the convoluted tubules. Estimates of the resultant proteinuria, in rats and in dogs, have enabled semi-quantitative assays to be made and potential antagonists to be tested. In FIGURE 1 is reproduced the relation between the logarithm of the dose of polymyxin A administered and the mean weight of protein excreted in 72 hours by a group of four rats (Wistar, inbred stock), expressed in terms of mg. per 100 gm. of rat. It is to be noted that the "normal" rat excreted 0 to 12 mg. protein. This test, which is capable of refinement, has proved useful.

It is known that the nephrotoxic action of *DL*-serine is due to the unnatural isomer, *D*-serine (Artom, Fishman, and Morehead⁸), and the site of damage to be the distal portion of the convoluted tubules (Wachstein⁹). Further, Wachstein⁷ found *DL*-methionine and glutathione, but not cysteine, thioglycolic acid or 2:3-dithiopropanol to protect against *D*-serine. *DL*-alanine, glycine, *DL*-threonine, and glycolic, butyric, and pyruvic acids gave some protection.

Brownlee and Short⁸ confirmed the effects of these substances in rats and dogs by estimating the degree of proteinuria, and have further found that substances which antagonize the nephrotoxic action of *D*-serine also antagonize the nephrotoxic activity of preparations of polymyxin A. A

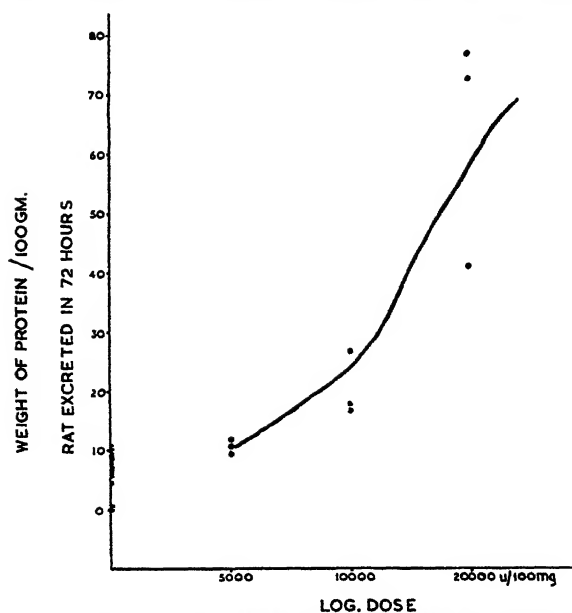


FIGURE 1 The semi-quantitative relation between logarithm of the dose of polymyxin A administered and the mean weight of protein excreted in 72 hours, by a group of 4 rats, expressed in terms of mg. per 100 gm. rat. The "normal" excretion is 0-12 mg.

number of the substances tested are listed in TABLE 7. Unless otherwise

TABLE 7
THE EFFECT OF VARIOUS SUBSTANCES ON THE NEPHROTOXIC FACTOR OF POLYMYXIN A
IN RATS

Dose 1 mg. (10, 000 units) per 100 gm. rat

Substance	Protein excreted in 72 hours, mg./100 gm. rat
Glycine, 400 mg.	24.5
<i>DL</i> -Alanine, 400 mg.	30.5
<i>L</i> -Cysteine, 400 mg.	40.0
<i>DL</i> -Methionine, 200 mg.	12.0
<i>S</i> -Methyl- <i>L</i> -Cysteine, 200 mg.	15.1
2:3-Dithiopropanol, 10 mg.	55.0
Protein hydrolysate horse muscle	
Parenterally, 100 mg.	17.4
Orally, 5 gms.	17.0
In diet, 50 mg. daily for 4 days	14.0
In diet, 1 gm. for one day	16.5
Polymyxin A alone	58.8, 57.5, 62.0
Controls	11.4, 12.3, 8.0

stated, they were injected parenterally. The most effective protecting agent is *DL*-methionine, followed by *S*-methylcysteine, then animal protein hydrolysate. Animal protein hydrolysate is also effective orally and, further, protection may be devised prophylactically by the latter means. One substance rich in reactive -SH groups, 2:3-dithiopropanol, is ineffec-

tive and *L*-cysteine gives little protection. Glycine and *DL*-alanine, which contain no sulfur, are intermediate in their protective capacity.

Protection by methionine is complete in the dog, as illustrated in TABLE

TABLE 8
THE PROTECTIVE EFFECT OF METHIONINE AGAINST THE NEPHROTOXIC FACTOR
IN DOGS

Methionine	Max. conc. protein in urine in 96 hours
Dose 0.8 mg. (8,000 units) per kg. 3-hourly 4 times	
<i>mg./kg.</i>	<i>gm./l.</i>
20	0
10	0
5	0
2.5	0, 0.2
1.0	0.8, 1.4
0.0	0.7, 0.2, 0.3
Dose 1.0 mg. (10,000 units) per kg. 3-hourly 4 times daily for 3 days	
10	0
5	0, 1.0
2.5	7.1
0.0	5.0, 2.0

8, where it is shown that 10 mg. of methionine injected once will protect against the nephrotoxic action associated with 1 mg. of polymyxin A given 4 times daily for 3 days. Complete protection with methionine is not seen in man. It is an interesting observation that the tubular damage in man appears to be quite temporary. This is illustrated in FIGURE 2, which

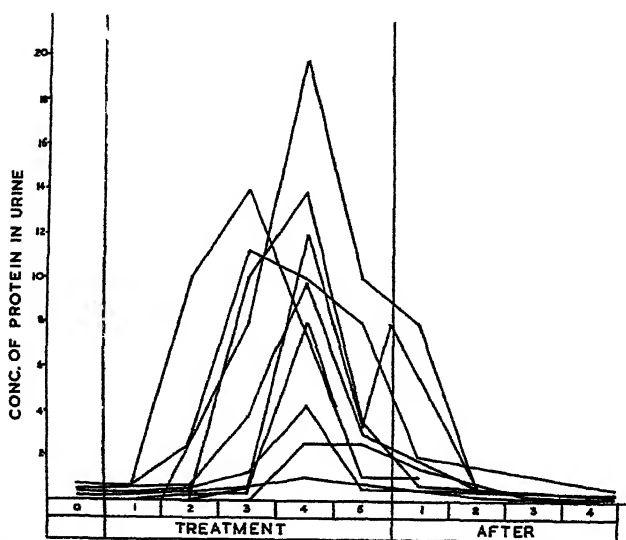


FIGURE 2. The concentration of protein in urine, in gm. per liter, related to the administration of Polymyxin A to man for 5 days. The stimulus is optimal at about the fourth day, and indicates that the kidney has mobilized repair processes by this time.

relates the concentration of protein in urine in gm. per liter to time in days. While the stimulus is applied for 5 days in the form of doses of polymyxin A, the response is skew: it is at a maximum at about the fourth day. In other words, the kidney has been able to mobilize its own repair devices during the period of exhibition of the damaging agent.

Serine is absent from toxic preparations of polymyxin A, and the hydrolytic product of the antibiotic and its individual amino-acid components (Catch and Jones⁹) leucine, threonine, and α,γ -diaminobutyric acid are not themselves nephrotoxic.

It may be deduced from these observations that the toxic principle is a peptide with an unnatural amino acid make-up, that the lesions arise during tubular resorption, and that labile methyl groupings may be involved in the process.

A Comparison of Kidney-Damaging Effects of Polymyxin A, B, and D in Rats

The total urinary protein excreted by groups of four Wistar rats in the 72 hours after the subcutaneous injection of 1 mg. per 100 gm. of a number of batches of polymyxin A was compared with that produced by a standard batch "A 1 P 48," and a control group receiving saline.

There has been listed, in TABLE 9, an index of the nephrotoxic effect of

TABLE 9
NEPHROTOXIC PRINCIPLE ASSOCIATED WITH POLYMYXIN A

Batch	Units/mg.	Wt. protein excreted mg.		
		Test	Standard	Control
First Hydrochloride				
2	5,250	164	335	55
12	3,400	300	129	9
23	5,612	323	235	52
24	4,800	540	235	52
36	6,000	180	298	48
37	4,000	438	209	44
85	5,000	146	107	27
104	6,500	247	130	37
105	5,900	204	130	37
106	5,500	176	130	37
Hydrochloride Prepared from Successive Helianthate Fractions				
256	8,750	148	280	75
257	5,330	234	280	75
258	8,320	235	280	75
259	6,960	135	280	75
260	2,430	236	280	75
Hydrochloride Prepared from Successive Reineckate Fractions				
54	7,940	214	238	49
55	8,260	116	238	49
56	8,500	139	238	49
57	7,460	234	238	49

ten representative plant-scale batches of polymyxin A. Following these are five hydrochloride fractions obtained by the successive helianthate fractions of one batch of the antibiotic. It is to be noted that there is no direct relation between impurity and nephrotoxic action. An alternative reineckate fractionation, also given, shows no diminution in nephrotoxic activity with purification. Even though preparations 256 and 56 are to be considered of only 90 per cent purity, it must be considered a possibility that the nephrotoxic property associated with polymyxin A is an intrinsic property.

TABLE 10
COMPARISON BETWEEN THE NEPHROTOXIC PROPERTIES ASSOCIATED WITH POLYMYXIN A, B, AND D IN THE RAT

Mean Weight of protein excreted per rat, in a group of four, in 72 hours				
Batch	Units mg.	Wt. protein excreted mg.		
		Test	Standard	Control
Polymyxin A, Hydrochloride Prepared from Successive Reineckate Fractions				
54	7,940	214	238	49
55	8,260	116	238	49
56	8,500	139	238	49
57	7,460	234	238	49
Polymyxin B, First Hydrochloride				
93	1,200	24	173	39
108	2,800	89	132	52
109	2,900	14	263	35
120	1,600	48	168	45
121	2,660	61	168	45
123	4,500	44	168	45
Polymyxin D				
	1,000	147	73	12
	1,000	210	260	35

* Stansly *et al* 2,000 u mg

When a comparison is made between polymyxin A, B, and D, it is observed that Polymyxin B causes no significant change in proteinuria in the rat, while polymyxin A and D cause a marked increase.

Renal Toxicity in Rabbits. (Comparison between polymyxin A and B). Rabbits given 3-hourly subcutaneous injections of 3 mg., 30,000 units per kg., four times daily for three days show no significant increase in proteinuria in the case of those given polymyxin B. Those treated with polymyxin A show a peak excretion of protein at 24 to 48 hours. A microscopical examination of the urine of rabbits given polymyxin B shows no casts and only an occasional renal cell, in contrast to those given polymyxin A, which showed the presence of renal cells and casts at the 24- and 48-hour periods, and renal cells at 48 and 72 hours.

Histological Examination. Rabbits numbered 1006 and 1010, receiving

TABLE 11
NEPHROTOXIC ACTION OF POLYMYXIN A AND B IN THE RABBIT TREATED WITH 30,000
UNITS PER KG 4 TIMES DAILY FOR 3 DAYS

Rabbit	Batch	u. mg.	Concentration of protein (g./l.) in urine after			
			0 Hrs.	24 Hrs.	48 Hrs.	72 Hrs.
Polymyxin A						
1006	52	4,340	0.18	0.3	1.2	1.8
1010	85	5,000	0	3.4	2.6	0.5
Polymyxin B						
1004	83	4,600	0.18	0.6	0	0.13
1008	83	4,600	0	0.2	0.02	0.08
1005	87	3,800	0.1	0.8	0.2	0.9
1009	88	4,500	0	0.14	0.27	0.24
1007	93	1,200	0	0	0.12	0.05
Controls						
			0	0	0.1	0.3
			0	0	0	0.08

polymyxin A, showed the usual congestion and stripping of the epithelium of many tubules with disintegration of cells of Henle's loops.

Rabbits 1004, 1008, 1005, 1009, and 1007, given polymyxin B, were normal or with slight cloudy swelling of some secretory tubules.

Comparison of Renal Toxicity of Polymyxin A and B in Dogs

A comparison of the nephrotoxic action of polymyxin A and B in dogs given 3-hourly subcutaneous injections of 1 mg./kg. four times a day for 3 days is shown in TABLE 12, in which typical results are given. This is a

TABLE 12
A COMPARISON OF THE NEPHROTOXIC ACTION OF POLYMYXIN A AND B IN DOGS
1 mg. (10,000 units) 4 times daily for 3 days

Dog	Batch	u /mg.	Concentration of protein in urine (g./l.) after					
			0 Hrs.	24 Hrs.	48 Hrs.	72 Hrs.	96 Hrs.	120 Hrs.
Polymyxin A								
8	37	4,000	0	0.1	2.6	4.4	0.8	—
11	52	4,340	0	0.08	3.7	1.0	0.5	0.25
18	85	5,000	0	0.8	2.5	1.4	—	—
14	85	5,000	0	0.15	1.8	2.6	1.0	1.0
Polymyxin B								
19	108	2,800	0	0	0	0.07	—	—
20	108	2,800	0	0.20	0.3	0.3	—	—
12	123	4,500	0	0.56	0.25	0.2	0.3	0.4
2	124	6,000	0	0.12	0.14	0.7	0.5	0.7
17	125	6,000	0	0.5	0.8	0.8	—	0.6

critical experiment in view of the known nephrotic tendency of the dog. As with the rat and rabbit, it is observed that polymyxin B differs from polymyxin A in causing no significant degree of damage. The peak excretion of protein is at 48 to 72 hours.

The microscopic examination of the urine of the treated dogs shows the presence of many renal cells but only occasional casts in animals treated with polymyxin A, and no or only occasional renal cells in the case of dogs treated with polymyxin B.

Histological Examination of Kidneys. Dogs treated with polymyxin A show severe acute nephrosis. Dog number 18 is a typical example (FIGURE 3). The secretory epithelium is seen in all stages of disintegration;

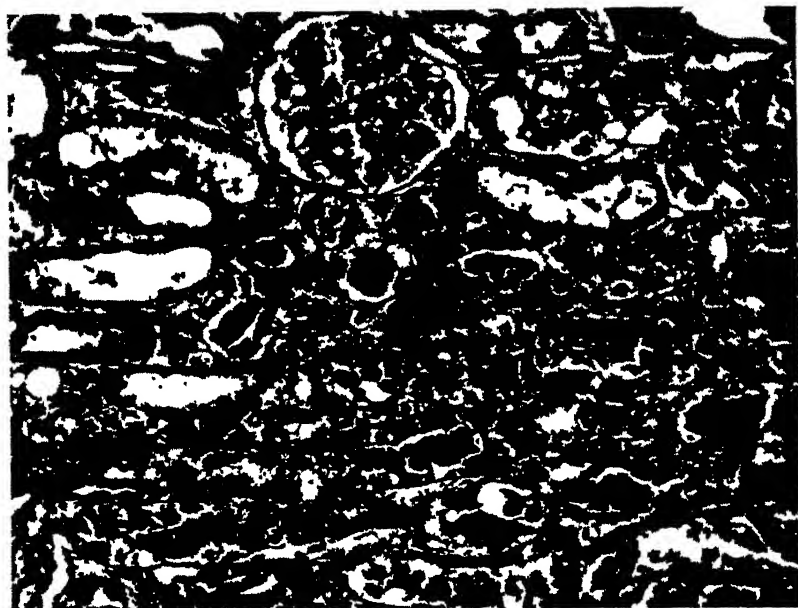


FIGURE 3. Kidney section from a dog (No. 18) treated with 10,000 units per kg. of polymyxin A 4 times daily for 3 days. X405 Stain Picromallory. There is severe acute nephrosis. The secretory epithelium is in all stages of disintegration; cells are shrunken, and there are large numbers of granular, hyaline, and cellular casts. There is no apparent glomerular damage apart from a slight swelling of tuft epithelium.

cells are shrunken rather than swollen; large numbers of granular, hyaline, and cellular casts are seen. There is no apparent glomerular damage apart from possibly slight swelling of tuft epithelium, which, however, is also seen in the "normal" dog. There is also seen a little active, acute and chronic, focal interstitial nephritis.

In dogs treated with polymyxin B, the only abnormalities observed in the histology of the kidney were a mild hyaline droplet degeneration of convoluted tubular epithelium and occasional foreign-body (round-worm larvae) granulomata (see FIGURE 4). Similar changes were observed in controls.

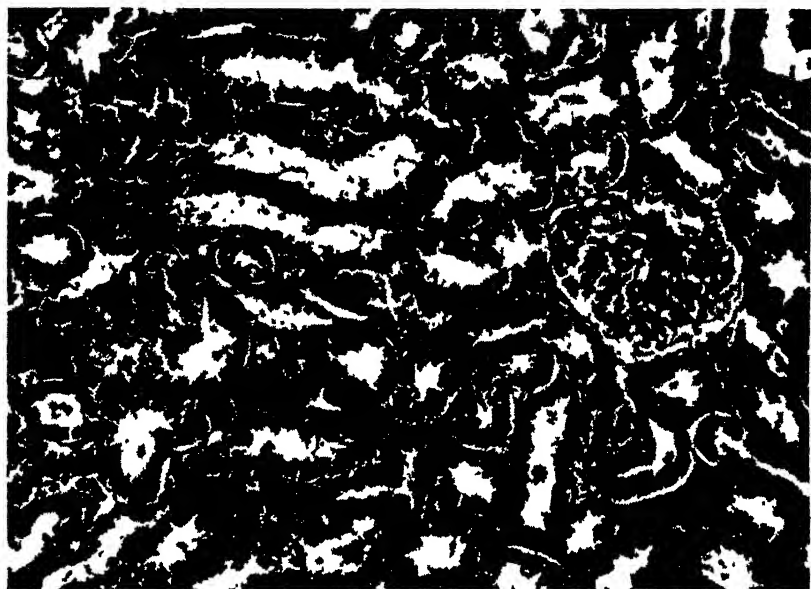


FIGURE 4. Kidney section from a dog (No 20) treated with 10,000 units per kg. of Polymyxin B 4 times daily for 3 days $\times 360$. Stain Picromallory. The only abnormality observed is a mild hyaline droplet degeneration of convoluted tubular epithelium. Similar changes were observed in the controls.

Renal Damage and Urinary Excretion of Polymyxin A and B in Dogs

Dogs given subcutaneous injections of 1 mg./kg. of polymyxin A excrete large amounts of the antibiotic in urine both during and for several days after treatment. The greater the kidney damage, the higher appears to be the concentration of antibiotic in the urine.

Similarly, the concentration of polymyxin B found in urine is small, but is adequate to sterilize gram-negative organisms (TABLE 13).

Nephrotoxic Factor

Enzyme Treatment. Neither the nephrotoxic principle nor the antibacterial activity of polymyxin A or D is affected by digestion with trypsin, pepsin or papain.

Renal Damage and Urinary Concentration in Man. A comparison is made, in TABLE 14, of the effects on urinary concentrations of both antibiotic and protein in man treated with polymyxin A and varying doses of methionine, and with polymyxin B.

As with the experimental animal, there appears to be a direct relation between the extent of renal damage and the concentration of antibiotic in the urine. Nevertheless, as we propose to demonstrate, the concentrations in urine are adequate to inhibit gram-negative pathogens.

Polymyxin B; Side Effects. Two apparently related side-effects have been observed after the administration of polymyxin B in experimental animals and in man. They are a mild pyrexia, reversible with antipyretic

TABLE 13

NEPHROTOXIC ACTION AND RENAL EXCRETION OF POLYMYXIN A AND B

"P" is the units in plasma "Uu" is the units in urine. "Up" is the protein in urine.
The damaged kidney (Poly myxin A) appears to leak antibiotic.

Poly- myxin	Dog	Poly myxin u. ml. Protein gm./l.	Day of treatment			Day after treatment		
			1	2	3	1	2	3
A	11	P.	8	8	32			
		Uu.	7	106	213	106	53	26
		Up.	0	0.1	3.7	0.7	0.5	0 25
	18	P.	4	16	8	4		
		Uu.	3	56	106	26		
		Up.	0	0.75	2.5	1.4		
B	12	P.	24	3	6			
		Uu.	3	3	213	106	2 6	7
		Up.	0	0.25	1 8	2 6	1 0	0 8
	8	P.	8	4	4			
		Uu.	3	3	3	3	3	3
		Up.	0	0	0 8	0 2	0	0
	19	P.	4	8	4	4		
		Uu.	3	13	26	13		
		Up.	0	0	0	0.07		

TABLE 14

A COMPARISON OF THE NEPHROTOXIC ACTIVITY OF POLYMYXIN A AND B IN MAN
No kidney damage is seen with Polymyxin B. As with the dog, the kidney damaged
with Polymyxin A appears to cause a leak of protein and antibiotic.

Poly- myxin	Case	Dose mg./ kg.	Methio- nine	Anti- biotic units, ml. Protein g l.	Day of treatment					Day after treatment		
					1	2	3	4	5	1	2	3
A	C	1	×5	u. ml.	—	—	250	380	73	<3	—	—
				g. l.	—	—	1 0	0.9	0.5	0.2	—	—
	E	1	×10	u. ml.	3	20	133	266	533	426	106	26
				g. l.	—	0 02	0 17	1.1	3.7	1.5	1 0	0 7
	Mac	1	×5	u./ml	6	13	7	13	106	326	106	26
				g. l.	—	—	0 2	0.2	1.4	4.4	0.7	0.3
	K	1	×10	u./ml.	7	26	53	26	213	326	213	26
				g l	—	—	0.1	0.8	0.6	1.4	0.5	0.1
	Bur	1	×5	u./ml.	—	3	3	7	106	213	106	13
				g./l.	—	—	—	Tr	2.5	1.7	2.0	Tr
B	B	0.5	—	u./ml.	3	7	<3	<3	<3	3	<3	<3
				g./l.	Tr	—	Tr	—	—	Tr	—	—
	W	0.2	—	u./ml.	3	3	<3	<3	<3	<3	<3	<3
				g./l.	—	—	—	Tr	Tr	Tr	—	—

drugs, and a mild local reaction which is of delayed onset. Both have proved to be of minor importance and have not proved contraindications to the use of the drug. The preparations in use are impure and it is not yet possible to indicate drug or impurity.

Summary

1. Polymyxin A, B, and D are powerful antibacterial substances selectively active against Gram-negative pathogens. They are chemotherapeutic.

In experimental infection due to *H. pertussis*, polymyxin A is superior to polymyxin B, and both are superior to polymyxin D. In the very acute infections with *E. typhosa* and *K. pneumoniae*, polymyxin A is more efficient than polymyxin B and D.

2. Estimates of toxicity to mice (L.D.₅₀), by intraperitoneal and intravenous routes, show polymyxin D to have one-half to one-third the acute toxicity of polymyxin A or B.

3. Polymyxin A and D causes gross renal damage in the rat, guinea-pig, rabbit, and dog. Polymyxin A causes proteinuria in man.

Polymyxin B produces slight proteinuria under experimental conditions but no histological changes in the kidney. It does not cause proteinuria in man.

4. The proteinuria may be prevented partly in the rat and totally in the dog by the simultaneous administration of methionine parenterally or orally, and by protein hydrolysate orally.

5. Polymyxin A and B appear in the urine. The concentration is greater with A than with B and in the former case is proportional to the damage to tubules.

6. Two side effects observed in experimental animals with samples of impure polymyxin B, a mild pyrexia reversible with antipyretic drugs and a local reaction, have not proved contraindications to the use of this antibiotic in man.

References

1. AINSWORTH, G. C., A. M. BROWN, & G. BROWNLEE. 1947. 'Aerosporin,' an antibiotic produced by *Bacillus aerosporus* Greer. *Nature* 160: 263.
2. BROWNLEE, G. & S. R. M. BUSHBY. 1947. Chemotherapy and pharmacology of Aerosporin. A selective Gram-negative antibiotic. *Lancet* 254: 127.
3. STANLEY, P. G., R. G. SHEPHERD, & H. J. WHITE. 1947. Polymyxin: a new chemotherapeutic agent. *Bull. Johns. Hopkins Hosp.* 81: 43.
4. KENDRICK, P. L., G. ELDERING, M. K. DIXON, & J. MISNER. 1947. *Am. J. Publ. Health* 37: 803.
5. ARTOM, C., W. H. FISHMAN, & R. P. MOREHEAD. 1945. The relative toxicity of *L*- and *DL*-serine in rats. *Proc. Soc. Exp. Biol. & Med.* 60: 284.
6. WACHSTEIN, M. 1947. Nephrotoxic action of *DL*-serine in the rat. I. *Arch. Path.* 43: 503.
7. WACHSTEIN, M. 1947. Nephrotoxic action of *DL*-serine in the rat. II. *Arch. Path.* 43: 515.
8. BROWNLEE, G. & E. I. SHORT. 1948. Antagonism by amino-acids of renal tubule damaging substances present in 'Aerosporin' preparations. *Biochem. J.* 42: lii.
9. CATCH, J. R., T. S. G. JONES, & S. WILKINSON. 1948. The Chemistry of Polymyxin A. *Ann. N. Y. Acad. Sci.* 51(5): 917-923.

POLYMYXIN: ASSAY PROCEDURES

By ELWYN T. REESE AND G. M. EISENBERG

J. T. Baker Chemical Company, Phillipsburg, New Jersey

At the Philadelphia meeting in 1947, we were much impressed by the report on polymyxin given by Dr. Benedict,¹ and with his cooperation we undertook a study of this promising antibiotic. The more the work progressed, the more interesting and difficult the problems became. It is our purpose to present some of the results of that work.

At that time, a great many procedures had yet to be worked out. The dilution assay in use required that the samples be sterilized prior to assaying. This necessitated the use of large numbers of bacterial filters, and was a time-consuming operation. Lacking the filters and being averse to slow filtrations, we turned to other means, as did Benedict at the same time. Apparently, both he and we considered the plate assay method, with discouraging results. Benedict then developed a turbidimetric method which is apparently satisfactory, but which he dropped on learning the details of the plate method developed at Cyanamid.

Our solution, though we were unaware of it at this time, was exactly the same as used by the English workers.² Recognizing that the test organism, *E. coli* (NRRL B-281), was penicillin-resistant and that our strain of *B. polymyxa* would probably be sensitive to penicillin, we set up tests to determine the critical penicillin concentrations. We found that with *B. polymyxa* but slight growth occurred at 0.8 μ ./ml. broth and none at 4 μ ./ml.; with *E. coli* (NRRL B-281) growth was good at 20 μ ./ml., slight at 50 μ ./ml.; and with *Brucella bronchiseptica* (NRRL B-140) growth good at 20 μ ./ml., none at 100 μ ./ml. It is apparent that a concentration between 4 and 20 units of penicillin per ml. will inhibit the development of the *B. polymyxa* in the assay broth and yet permit the development of the assay organism, be it *E. coli* or *Brucella bronchiseptica*.

The dilution assay procedure we now employ follows:

To one liter of Difco Dextrose broth, add 10 ml. of 1 per cent phosphate buffer pH 6.8, 10 ml. of stock penicillin solution (1000 μ ./ml. in buffer), and 1 ml. of a 16-hour culture of *E. coli* in Trypticase soy broth. Pipette into sterile but unplugged test tubes, varying the amount where desired. One ml. of unknown is added to the first tube and successive transfers of 1 ml. made to the remaining tubes. The tubes are covered and incubated at 37°C. overnight. The end point is taken as the last clear tube. Each unknown is run in triplicate.

At the time the above procedure was developed, many interfering factors were unsuspected. Even now, we wish we knew how to explain some of the discrepancies that occur. We found that a different brand of dextrose broth was not suitable, the growth of the organism being very poor. We found further, that, when we sent one of our samples to other laboratories for assay, the results were much different from our own (TABLE 1).

Recognizing the possible effect of medium, we tried using Trypticase soy broth as our diluting medium, and obtained the same results as Cyanamid,

TABLE 1
EVALUATION OF POLYMYXIN SAMPLE BY DILUTION TECHNIQUE IN VARIOUS
LABORATORIES

	Inhibiting concentration of P-11-DAXM vs. <i>E. coli</i> *
Baker assay (Difco broth)	.08 µg./ml.
Cyanamid assay (Trypticase soy broth)	1.1-1.5 µg./ml.
Peoria assay (Trypticase soy broth)†	.32
Baker assay (Trypticase soy broth)	1.0-1.2

* *E. coli* NRRL B-281 used by all but Cyanamid, who used *E. coli* (McCloud strain).

† Peoria assay broth contains glycine buffer.

with whose assay technique ours was now identical. The concentration of polymyxin required to inhibit growth of *E. coli* thus appears to be a function of the medium. Several tests were run to confirm this point (TABLE 2).

TABLE 2
COMPARISON OF POLYMYXIN ASSAY RESULTS ON DIFFERENT ASSAY BROTHS

Test	Penicillin (10 µ/ml.)	Inhibitory conc. of polymyxin* (µg./ml. broth)	
		Difco dextrose broth + buffer	Trypticase soy broth
I	+	0.045	0.9
II (a)	—	0.5	2.0
(b)	+	0.12	0.5
III (a)	—	0.04	0.3
(b)	+	0.04	0.08

* *E. coli* as test organism.

There can be no questioning the fact that the amount of polymyxin required for inhibition is less in Difco Dextrose broth than in Trypticase soy broth. This appears to be correlated with improved growth of the test organism in the latter medium, and may be a function of the nature of the proteinaceous ingredients of the different products.

A second factor affecting the results is the presence of penicillin in the assay broth (TABLE 2 and 3). It seems that the concentration of penicillin

TABLE 3
EFFECT OF PENICILLIN ON POLYMYXIN ASSAY

Penicillin* (µ/ml. broth)	Inhibitory conc. polymyxin (µg./ml. broth)	
	vs. <i>E. coli</i>	vs. <i>Br. bronchiseptica</i>
0	3.3	.33
1.	1.1	.33
3.	—	.33
10.	1.1	.11
20.	.12	—
30.	—	.05
50	No growth of <i>E. coli</i>	—

* Amorphous penicillin, J. T. Baker Chemical Co.

which we chose to use is a borderline one, the result being that a small change in penicillin concentration might give a large difference in assay value. This would, in a large measure, account for day-to-day variation in our assay results. It would appear that lower concentrations of penicillin—about 5–7 μ . ml.—should be more satisfactory for the regular assay procedure.

It is interesting to speculate as to why the concentration of polymyxin required to inhibit is lowered by the presence of penicillin. A similar effect was noted between streptothricin and penicillin by Foster and Woodruff,⁴ but such a phenomenon could not be shown for gramicidin and penicillin by Herrell and Heilman.⁵ Foster and Woodruff reached the conclusion that the action was additive and that, therefore, both antibiotics acted on a common mechanism. There are many explanations possible which do not assume action on a common mechanism, namely, that the penicillin may assist the polymyxin by any one of the following means:

(a) Interference with the enzymatic inactivation of polymyxin, thereby maintaining a toxic level for a longer time.

(b) Interference with other systems capable of blocking the active group of polymyxin, or of tying up the polymyxin in an insoluble complex.

(c) Indirectly by affecting the permeability of the cell membrane, thereby permitting the polymyxin more rapid access to the cell protoplasm.

By any of these means, the combined action might be either additive or synergistic. In the case of polymyxin and penicillin, total inhibition was used as the end-point, and it is not possible for us to say whether the total action is more than additive.

The dilution assay, at times, has contained a clear tube interrupting a turbid-tube series or a turbid tube in a clear tube set. When such an aberrant tube occurs near the polymyxin end-point, errors are likely to arise, and greater reliance must be placed on the replicate series. Of the possible sources of such trouble, the detergent used in cleaning the tubes was selected as worthy of attention, since traces of a white substance were evident in the bottom of some tubes. Attention was first directed to the effect of the Alconox on the assay organisms. It was found that 0.5 per cent Alconox in broth will inhibit *E. coli* and that less than 0.05 per cent will inhibit *Brucella bronchiseptica*. A clear tube in a turbid series could thus result from concentrations of Alconox in excess of those shown.

The effect of Alconox in the broth on activity of polymyxin was determined against *E. coli* (TABLE 4). The addition of as little as 0.01 per cent

TABLE 4
EFFECT OF ALCONOX ON INHIBITION OF *E. coli* BY POLYMYXIN

	Inhibitory concentration μ g./ml.	
	Test I	Test II
Broth	0.37	0.04
Alconox 0.1% in broth	>4.0	11.0
Alconox 0.01% in broth	—	.4

Alconox raised the minimum inhibiting concentration at least ten-fold. It is believed that this inactivation results from complex-formation between the detergent and antibiotic. A turbid tube in a clear series might easily be due to such an inactivation of the polymyxin, the amount of Alconox being sufficient to reduce the active polymyxin concentration, but too low to inhibit growth of the bacterium.

The amount of inoculum used in the dilution assay has a marked effect on the results obtained (TABLE 5). Such an effect has been observed for

TABLE 5
EFFECT OF AMOUNT OF INOCULUM ON POLYMYXIN ASSAY VALUES

	Inoculum in drops/100 ml. broth	Inhibitory conc. polymyxin (μ g./ml. broth)	
		vs. <i>E. coli</i>	vs. <i>B. bron-</i> <i>chiseptica</i>
Difco broth	2	0.7	0.013
Difco broth	10	6.0	0.37
Difco broth	50	17.0	1.1
Trypticase soy broth	2	2.0	0.1
Trypticase soy broth	40 (1.5 ml)	17.0	3.3

each of our test organisms, and for both dilution media. The greatest variation obtained in this test was a 30-fold increase in the concentration necessary to inhibit *B. bronchiseptica* resulting from a 5-fold increase in inoculum. The effect is greatest for the weaker inocula, the increase in polymyxin required to inhibit being less for the larger number of bacteria. Since no bacterial counts were made, there is no way of knowing the relation between the numbers of bacteria in the *E. coli* and in the *B. bronchiseptica* tubes. Were this known, it would be possible to compare the susceptibility of the two test organisms. For, although on the surface it appears that *B. bronchiseptica* is the most sensitive organism we have tested, the numbers may be very low and the sensitivity not real. For example, if in our experiment the numbers of cells in the *Brucella* tubes is only $\frac{1}{10}$ that in the *E. coli* tubes, then both organisms have the same sensitivity to polymyxin. The importance of this is obvious. If we run a spectrum for polymyxin, the results will be a function of the numbers of each organism employed. The numbers, further, will be related to the cultural conditions under which the inoculum is prepared. It will be small wonder, therefore, if two laboratories arrive at different answers. Indeed, our own ratios, run at different times, vary considerably, and we find that the only comparisons that have any value are those between samples run simultaneously.

The only other factor we have examined by this means has been the effect of horse serum on the inhibiting concentration. Our results indicate that at least 5 times as much polymyxin is required for inhibition of *E. coli* in dextrose broth containing 10 per cent horse serum, as is required in broth lacking the serum.

Comparison of Results by Dilution and by Plate Assay Methods

The plate assay method we have used is that developed by the Cyanamid⁷ group, as modified by Benedict.² We use 0.5–1.0 ml. of a 20 hour culture of *Brucella bronchiseptica* per 100 ml. of seed agar containing Tween 80. The surface of the agar is dried for 15–20 minutes at 37°C. by partially removing the lids of the petri plates prior to adding the half-inch discs containing the polymyxin samples. At the end of 16–20 hours' incubation, large, sharply defined zones are formed which are easy to read. All unknowns are diluted with glycine-HCl buffer (pH 2.0–2.2) to fall in the range 33–100 Cyanamid units per ml. Each plate contains a 100 μ and a 33 μ Peoria standard polymyxin, and two unknowns. This procedure has proven so satisfactory that we have made no effort to improve on it. Some of the factors affecting the dilution assay results were tested for their effect on polymyxin values by the plate assay procedure. Alconox at a 0.5 per cent level, gelatin at 0.25 per cent, and horse serum at 50 per cent, each effected an apparent loss of about 25 per cent in potency. This 25 per cent loss is to be considered a maximum value, since the pH values rose a few tenths of a pH unit in some cases. There is, thus, no correlation between the effects of these materials on the different assay methods, though it must be recognized that the pH factor is radically at variance. Penicillin, at 25 μ per ml., a condition which promoted the action of polymyxin in the dilution assay, resulted in an insignificant 8 per cent loss by the plate method. Again, it must be realized that, at the low pH in the latter procedure, penicillin is rapidly inactivated.

Fortunately, or unfortunately, the above plate assay procedure did not become available until our work had been in progress for many months. Otherwise, we should never have been called on to evaluate the information obtained by one method in terms of the other. One of our first problems was the development of a fermentation medium giving maximum potency, as determined by the dilution assay method. It can be seen (TABLE 6)

TABLE 6
STRAINS OF *B. polymyxa* VERSUS MEDIUM

Strain Baker No.	Maximum μ /ml.			
	Dilution assay (<i>E. coli</i>)		Plate assay (<i>Brucella</i>)	
	Baker medium	B. St.*	Baker med.	B. St.*
Exp. I B-116	75,000	500	60	16
B-117	55,000	1000	52	46
B-118	75,000	1000	46	38
B-119	55,000	500	50	46
Exp. II B-116	50,000	500	72	31
B-117	5,000	500	74	45
B-118	10,000	100	124	21
B-119	10,000	400	148	22
Exp. III B-116	30,000	—	66	—
B-143 (NRRL B-694)	30,000	—	276	—

* Stansly medium plus peanut meal and chalk as recommended by Benedict.

that a 100-fold increase in potency (dilution assay) has been achieved as a result of changes in the medium, but without a corresponding increase in plate units. These results are based on shaker-flask tests, assays being made on the 3rd, 4th, and 5th days. Strain differences were not appreciable except in the third experiment, in which we compared our strain developed for dilution units against the NRRL strains developed for plate units. In this case, the strains were equal in dilution units, but the NRRL strain is much superior in plate units. If the ratios of dilution units to plate units are considered, it will be seen that the ratio differs (1) with the medium used, and (2) with the strain. The lack of correlation between the two types of units is even more evident when we follow the course of a fermentation (FIGURE 1.) In this experiment, three different conditions

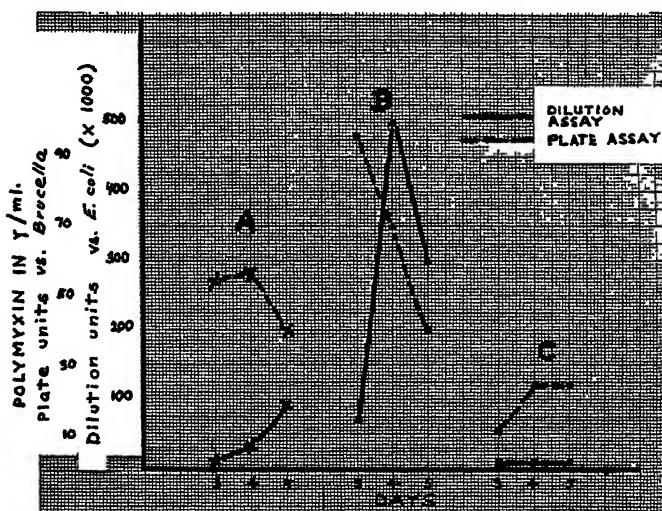


FIGURE 1. Polymyxin: dilution vs. plate assay.

(N. B. Legend should read "Polymyxin in μ /ml.")

were compared by two assay methods. In the earliest stages of the fermentation, both types of units must have developed simultaneously. After reaching a peak in A and B, the plate unit potency fell off while the dilution units continued to increase. In C, an increase in plate units was not paralleled by any increase in dilution units. Such a lack of correlation has frequently been encountered.

A comparison between dilution units and plate units on material at different stages in the process (TABLE 7) shows that the magnitude of the difference is greatest in assays of the broth. The lyophilized eluates or acetone-precipitated materials have a much lower ratio of dilution to plate units. In most cases, considerable variation occurs between batches. The evidence from several angles, therefore, is conclusive that there is no correlation between the two types of units, and the deduction follows that the selection of an assay method predetermines the interpretation of the results.

TABLE 7
COMPARISON OF DILUTION AND PLATE ASSAYS AT VARIOUS STAGES IN PROCESSING

	Dilution Assay	Plate Assay	Ratio Dilution Plate
Broths (Shaker) 1	5,000 μ ml	82 μ /ml	60:1
(2)	25,000	73	350:1
(3)	55,000	93	600:1
(4)	400,000	159	2500:1
Dried cultures 48 E L	1,800 μ mg	33 μ /mg	60:1
48 F L	420	27	15:1
48 G L	530	20	25:1
Acetone ppt 45 E L P	830 μ , mg	65 μ /mg	12:1
47 E L P	1,750	73	25:1
21 E L P	2,300	84	30:1
31 E L P	2,100	126	16:1

Variation in Slope of Plate Assay Curves

For some time, it has been observed that the concentration of our polymyxin samples frequently has an effect on the values obtained by the plate assay method (TABLE 8). The greater the dilution of the sample,

TABLE 8
EFFECT OF CONCENTRATION OF SAMPLE ON RESULTS BY PLATE ASSAY METHOD

Sample	Conc mg/ml	μ mg Value	Sample	Conc mg/ml	μ /mg Value
P 11 MC	10	35	P 11 DAXM	1	86
	5	38	(done by R Benedict)	0.5	85
	2.5	42		0.2	86
	1.25	51		0.1	90
P 10 AB	10	31	P 11 DAXM	3.0	55
	5	39	(done at Bacter Chem Co.)	1.5	57
	2.5	56		.75	60
	1.25	66		.38	80
P 14 E ₂	3.0	40	P 15 A	3.0	38
	1.5	57		1.5	40
	0.75	69		.75	40
	0.38	88		.38	41

the higher have been the values obtained. The pH has been checked in these cases, but did not vary more than 0.1 pH unit from that of the Peoria standard. Not all our samples show this trend, and those that do may vary in the absolute amount by which the lowest and highest values differ. Another way of expressing this result is in terms of the slopes of the curves obtained (FIGURE 2). Those substances giving increasing values for lower concentrations of material give a curve having a steeper slope than that of the Peoria standard. Two explanations are possible. The sample may contain impurities of such a nature as to interfere with the diffusion of the

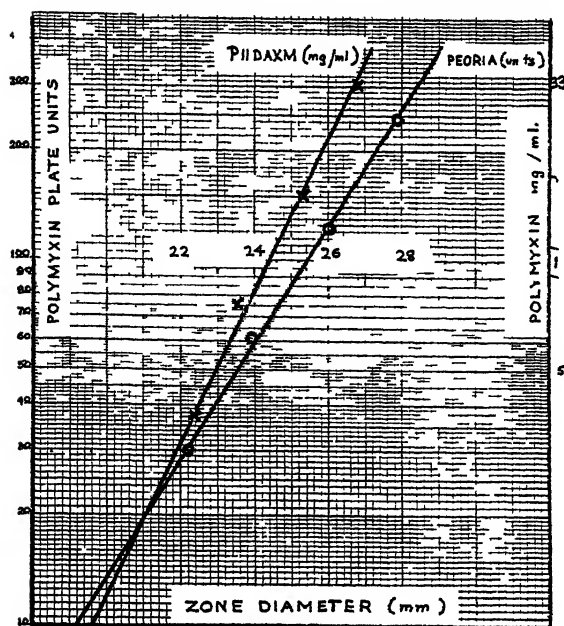


FIGURE 2 Polymyxin plate assay vs. *Brucella bronchiseptica* (semi log)

polymyxin. Apparently, conditions must be made suitable for polymyxin diffusion as evidenced by the necessity for Tween 80 and for low-pH samples. Given these conditions, however, no factors have been recognized as interfering with the free movement of polymyxin through agar.

The other explanation requires that polymyxin be a mixture, or that there be more than one polymyxin, the one with which we are here concerned being different from the Peoria standard. Other factors remaining constant (medium, test organism, cultural conditions), solutions giving curves having different slopes (K) contain different antibiotics. The major factors determining the slope are the diffusion rate (D) and degree of toxicity (T) of the antibiotic to the test organism. For two antibiotics, five possibilities exist:

- (1) $D_1 = D_2, T_1 = T_2$ then $K_1 = K_2$
- (2) $D_1 = D_2, T_1 > T_2$ $K_2 > K_1$
- (3) $D_1 > D_2, T_1 = T_2$ $K_2 > K_1$
- (4) $D_1 > D_2, T_1 > T_2$ (all combinations)
- (5) $D_1 > D_2, T_1 < T_2$ (all combinations).

It is possible that our active molecules are larger than those of the Peoria polymyxin, thereby accounting for a lower diffusion rate, or perhaps they have a lower degree of toxicity to the bacteria. We prefer to believe that mixtures of antibiotics are present, though such a preference is not based alone on differences in the slopes of the assay curves. From a theoretical mixture of two independent antibiotics, it would be possible to obtain the following results:

(a) No zone formation—both antibiotics either of low toxicity to the test organism, or else non-diffusible.

(b) A straight-line curve over the range examined, in which case the activity of only one component would be apparent.

(c) A jointed curve due to the intersection of two straight-line curves.

Variation in Ratio of E. coli Units to Br. bronchiseptica Units

Recognizing the limitations imposed by medium and bacterial numbers on a bacterial spectrum determination, we have nevertheless thought it advisable to compare two of our samples with samples from other sources (TABLE 9). Peoria 980 is a sample from Dr. Benedict containing 980

TABLE 9
BACTERIAL SPECTRUM

	Inhibitory conc. γ /ml.*			
	P 11 DAXM	P14E ₁₂	Peoria 980	Cyana mid 180
B 9 <i>B. subtilis</i>	8	12	12	25
14 <i>S. aureus</i>	50	50	25	>50
29 <i>Aerobacter aerogenes</i>	2	1.5	.75	3.
45 <i>Serratia marcescens</i>	>50	>50	>50	>50
64 <i>Salmonella enteritidis</i>	>1.	>1.	0.5	>1.
138 <i>E. coli</i> (McCloud)	3.	3.	0.1	1.
119 <i>E. coli</i> NRRL B-281	3.	1.	.33	1.
120 <i>Brucella bronchiseptica</i> NRRL B-140	.01	.01	<.0003	.1
123 <i>S. typhimurium</i>	3.	>10.	3.	>10.
124 <i>K. pneumoniae</i>	3.	3.	3.	3.
125 <i>S. paratyphi</i>	3.	3.	.11	1.
127 <i>S. schottmuelleri</i>	10.	3.	3.	10.
130 <i>Shigella dysenteriae</i>	>1.	>1.	0.5	>1.
139 <i>B. cereus</i>	11.	11.	0.3	4.
140 <i>Ps. aeruginosa</i>	11.	11.	11.	33.

* Trypticase soy broth as assay medium.

Cyanamid plate units per mg.; Cyanamid 180, a sample from Dr. Stansly containing 180 plate units per mg.; and our samples P11DAXM 68 μ /mg., and P14E₁₂ 76 μ /mg. Trypticase soy broth was used as the assay broth to conform with the procedures of others. As a result, the inhibiting concentrations are much higher than those previously found by us in Difco Dextrose broth, and using the same polymyxin sample.

Obviously, the four samples used are quite different based on Cyanamid plate units. When, however, they are compared weight for weight by the dilution method, the substances appear to have very nearly the same inhibitory power against many of the test organisms. In general, we can say that all four polymyxin samples are least active against *Serratia marcescens* and *Staphylococcus aureus*, moderately active against *E. coli* and most of the others, and highly active against *B. bronchiseptica*.

An interesting difference appears in the response of *E. coli* and of *B. bronchiseptica* to the different polymyxin samples. It appears that our samples are slightly less active against *E. coli* than is the Cyanamid 180. Against *Brucella*, the reverse is true, our samples being much more active. The greatest differential between the amount required to inhibit *E. coli* to the amount required to inhibit *Brucella* is demonstrated by Peoria 980, and is of the magnitude of one thousand times. Other differences in response are apparent from the table, but only the *E. coli-Brucella* case has been subject to repeated checking.

A group of our samples was compared with the Peoria standard (TABLE 10). The figures shown are averages of triplicate determinations

TABLE 10
RELATIVE TOXICITY OF POLYMYXIN TO *E. coli* AND TO *Brucella bronchiseptica*

Sample	Plate assay (<i>Brucella</i>) μ./mg.	<i>E. coli</i>			<i>Brucella</i>		
		μg./ml. Required to inhibit*					
		Exp. I	Exp. II	Avg.	Exp. III	Exp. IV	Avg.
35 FGLP	36	3.3	3.7	3.5	1.0	1.0	1.0
P 11 DAXM	68	.6	.2	.4	.05	.02	.035
P 11 DAM	71	.37	.12	.25	.004	.01	.007
40 ELP	71	.12	.12	.12	.03	.01	.02
28 GLP	110	.04	.01	.025	.0012	.0014	.0013
Peoria	980	.14	.02	.08	.00013	.0006	.00036

* Difco Dextrose broth assay medium.

for the dilution assays. The assays were performed on four different days, all samples being assayed against one organism on each day. The results are not in what we would like to call agreement, but they are typical of the variation obtained in day-to-day tests. Furthermore, the relationships between samples holds up fairly well.

In continuation of our search for differences, may we point out that our sample 28GLP is more potent than the Peoria material against *E. coli*, yet its activity against *Brucella* is less than that of Peoria 980 in both the plate and dilution assays. In terms of *E.c./Br.b.* ratios, our samples run from 3 to 40, while that of the Peoria 980 is 200. Again, we wish to emphasize that the actual values of the ratio are not constant, but that for any pair of samples assayed simultaneously the relationship of the ratios is in the same direction, that is, the ratio for the Peoria sample is always greater than the ratio for 28GLP.

We would be remiss if we failed to call attention to similarities between samples. In a general way, there is a correlation between the potencies determined by the plate method and by the dilution assays, those samples at the top of the table (TABLE 10) being much less toxic than are those at the bottom. A second similarity is that all substances are more toxic to *Brucella* than to *E. coli*.

Discussion

While recognizing the possibility of several interpretations for the results presented, we choose to develop only one, namely: that mixtures of polymyxins are produced by *B. polymyxa*. It is our belief that the polymyxins represent a class of compounds having some common features, such as ease of adsorption on carbon, solubility, and heat stability, and differing in other respects, one of the recognizable differences being that of size. In the light of this, suppose we consider two hypothetical polymyxins: Px being a comparatively diffusible type, and Pn much less rapidly diffusible, each varying a little in specific toxicity from the other. Our work began with a strain low in Px and Pn, but by developing a medium for that particular strain we were able to favor an increase in Pn with substantially no increase in Px (i.e., a medium giving very high dilution units and low plate units). Concurrently, Benedict³ developed a strain for a particular medium which favored the development of Px.

During the course of a fermentation both types of units develop, but under certain conditions Px may be converted into Pn, thereby resulting in an increase in Pn at the expense of Px, as seen in FIGURE 1. During the purification process, both polymyxins are adsorbed readily on carbon, but during the elution step less of the Pn is released. This would account for the decrease in the dilution/plate ratio so frequently observed. If purification is continued along the lines developed for Px, then Pn decreases and the dilution/plate ratio approaches a value which is constant for Px.

Another example of the purification of a mixture indicates a separation of components. Our sample P15 is a precipitate from a methanol solution by excess of acetone. When this precipitate was further treated with a small volume of methanol, two fractions were obtained: 15A, the undissolved fraction; and 15AB, that which dissolved and was subsequently dried. Assays of these two substances show sharp differences. Both were identical with respect to their inhibition of *E. coli*, the inhibiting concentration being about 0.37 $\mu\text{g./ml.}$ Difco Dextrose broth. The response to *Br. bronchiseptica* was quite different, the fraction taken up in methanol (15AB) being much more toxic than the residue (P15A) to this organism. The actual figures are:

P15A	38 Cyanamid plate units/mg.; 2,000 dilution units/mg.
P15AB	96 Cyanamid plate units/mg.; 50,000 dilution unit/mg.

It is unfortunate that more of our attention has not been directed to obtaining homogeneous samples of each polymyxin type. But it has seemed to us that many of the fundamentals have yet to be worked out, especially with respect to the dilution assay, and it has been our feeling that without such adequate methods it is impossible successfully to carry through a chemical separation of closely related polymyxins.

Summary

The results may be summarized as follows:

- (1) Mixtures of polymyxins are produced by some strains of *B. polymyxa* (at least the one with which we are working).

(2) Such mixtures are detectable biologically by (a) differences in values obtained by plate and by dilution assay procedures, (b) difference in response to two test organisms *E. coli* and *Br. bronchiseptica*, and (c) perhaps by the slope of the curve in the plate assay procedure.

(3) The entire course which an investigation of polymyxin shall take will be dependent upon the assay procedure used. Modification in medium or strain will yield a predominance of one polymyxin if the plate assay method is used, and a predominance of a different polymyxin if the dilution assay is employed. Similarly, chemical purification will result in obtaining one or the other of a mixture depending on which assay method is used.

Bibliography

1. BENEDICT, R. 1947. Report on Polymyxin presented at S. A. B. meetings in Philadelphia, May 1947.
2. BENEDICT, R. 1947-1948. Personal communications.
3. BROWNLEE, G. & S. R. M. BUSHBY. 1948. Chemotherapy and pharmacology of Aerosporin. *Lancet* 1: 127-132.
4. FOSTER, J. W. & H. B. WOODRUFF. 1943. Microbiological aspects of streptothricin. II Antibiotic activity of streptothricin. *Arch. Biochem.* 3: 241-255.
5. HERRELL, W. E. & D. H. HILMAN. 1943. Cytotoxic and antibacterial activity of gramicidin and penicillin. *Am. J. Med. Sci.* 206: 221-226.
6. STANSLY, P. G. 1948. Personal communications.
7. STANSLY, P. G. & M. E. SCHLOSSER. 1947. Studies on polymyxin: an agar diffusion method of assay. *J. Bact.* 54: 585-597.

A SIMPLE METHOD FOR THE ASSAY OF POLYMYXIN IN BLOOD AND URINE*

By P. G. STANSLY

*Chemotherapy Division, Stamford Research Laboratories, American Cyanamid Company,
Stamford, Connecticut*

Extension of the agar diffusion method previously reported¹ for the assay of polymyxin in fermentation liquors and concentrates has resulted in a rapid, simple, and relatively accurate method for the determination of polymyxin in blood and urine.

Seeded plates are prepared as previously described.¹ The test organism, however, is *Brucella bronchiseptica* N.R.R.L. B-140² and the completed plates are incubated at one temperature, namely 37°C., overnight. In order to economize on blood, $\frac{1}{4}$ inch filter paper discs (Schleicher & Schuell #740E) instead of $\frac{1}{2}$ inch discs are used. These discs require 0.03 ml. of solution to saturate.

The determination of polymyxin in blood requires a standard blood assay curve. A series of standard solutions of polymyxin in 50 per cent "acid blood" is prepared and a straight line constructed relating zone diameter to log (concentration of polymyxin \times 2). The acid blood is prepared by mixing equal volumes of whole blood and 0.2N HCl. Similarly, blood to be assayed is diluted with an equal volume of 0.2N HCl. The concentration of polymyxin in micrograms per ml. of whole blood is then read off the standard curve. A typical standard curve is illustrated in FIGURE 1.

It is recommended that blood which is to be assayed or used for the preparation of the standard solutions either be untreated, with respect to anticoagulants, or citrated, but not oxalated. Oxalated blood frequently results in a viscous solution upon the addition of 0.2N HCl.

Advantages of this method of assay are as follows: (1) 0.25 micrograms of polymyxin per ml. of blood may be determined in triplicate with a 0.05 ml. sample of blood; (2) solutions of 50 per cent acid blood appear to remain fluid indefinitely and are stable with respect to the antibiotic; (3) the pH of the blood solution (generally around 3.75) and the nature of the assay method preclude the necessity for strict aseptic precautions; (4) in the absence of antibiotic, inhibition zones are not produced by normal dog, mouse, or human blood treated as described.

The determination of polymyxin in urine also requires a standard curve, which in this case is simply obtained from standard solutions of polymyxin in glycine-hydrochloric acid buffer of pH 2. The sample of urine to be assayed is diluted with an equal volume of glycine-hydrochloric acid of pH 1.4, which generally results, in the case of dog urine, in a pH of 2.

Ten replicate determinations on a sample of blood obtained from a dog treated with polymyxin gave a coefficient of variation of 7 per cent on one

* A more detailed account of this method has since appeared in the Proceedings for the Society of Experimental Biology and Medicine. 68: 301, 1948.

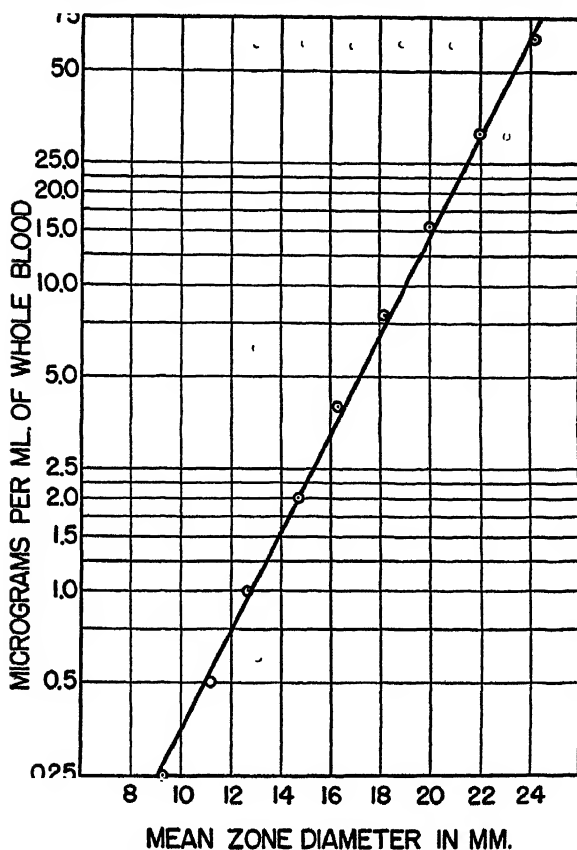


FIGURE 1. Standard curve for assay of polymyxin in blood.

day and 10 per cent when repeated on another day. The coefficient of variation of the twenty determinations was 12 per cent.

Bibliography

1. STANSLY, P. G. & M. E. SCHLOSSER. 1947. Studies on polymyxin: An agar diffusion method of assay. *J. Bact.* 54: 585.
2. BENEDICT, R. G. & F. H. STODOLA. 1948. Modification of an agar diffusion method of assay for polymyxin. *J. Bact.* 55: 286.

THE EFFECT OF POLYMYXIN IN THE TREATMENT OF EXPERIMENTAL BRUCELLOSIS AND TULAREMIA

By C. L. LARSON, B. N. CARLE, AND A. E. VERDER

National Institute of Health, Bethesda, Maryland

Recently, antibiotic substances have been isolated from *Bacillus polymyxa*. In the United States, particular attention has been given to polymyxin^{1, 2} and in England to Aerosporin.^{3, 4, 5} The need for effective therapeutic agents in brucellosis prompted a study of polymyxin to determine whether this substance could be employed in the treatment of *Brucella suis* infections in guinea pigs. The effect of polymyxin in tularemic infection of white mice was also studied. The data obtained indicate that polymyxin has no value in the treatment of experimental brucellosis in guinea pigs or of tularemia in white mice.

Materials and Methods

The organisms employed were *Brucella suis* strains 6 and 18, which were isolated in Indiana from human cases of brucellosis, by Dr. S. Damon of the Indiana State Board of Health. The strains are smooth and virulent for guinea pigs.

Polymyxin was obtained from two sources. The samples of polymyxin tested in the first two experiments were obtained from the Northern Regional Research Laboratory, Peoria, Illinois and consisted of two lots assaying 1,000 and 900-950 units per mg. The material tested in the third experiment was prepared by J. T. Baker Chemical Company, Phillipsburg, New Jersey and assayed 6,000 units per milligram. The drug was administered subcutaneously or intramuscularly after solution in saline or distilled water.

Guinea pigs of either sex weighing about 350 grams were selected. The animals were inoculated subcutaneously or intraperitoneally with the desired number of organisms. *Br. suis* was grown on Trypticase soy agar and incubated at 37°C. for 24 hours. The organisms were harvested, suspended in Zobell's solution, and centrifuged to remove any agar. The number of organisms was determined by observation of the density of the suspension, employing a Fischer photoelectrometer, and checked by replicate plate counts of appropriate dilutions. After an incubation period of 10 days or more, the animals were weighed and treatment was started and continued for 10 to 14 days. In two experiments, post-treatment periods of 1 to 2 weeks elapsed before the animals were weighed, bled, and killed. The pathological lesions were observed, the spleen weighed, and cultures on Trypticase soy media were made of the liver and spleen in all instances and from the blood and kidney in some. Agglutination titers were determined on the serums of all animals, employing the technique recommended by the Bureau of Animal Industry.

Only one experiment was made to determine the effect of polymyxin

upon tularemic infections in white mice. White Swiss mice of either sex weighing 20 grams were infected with a suspension of a virulent strain of *Pasteurella tularensis*. Twenty-four hours after administration of the infective dose of organisms, treatment was instituted and continued until all animals succumbed.

Results

Experiment 1. A group of 24 guinea pigs weighing about 350 grams was divided into 2 lots one of which was infected with 12 and the other with 12,000 *Br. suis* organisms strain 18 intraperitoneally. Each lot was then further divided to contain 6 untreated control animals, 3 animals treated subcutaneously with 6.4 mg. of polymyxin (lot 28A containing 1,000 u./mg.) per kg. of body weight once daily, and 3 animals treated twice daily in a similar manner with the same amount of polymyxin. The infective dose was given on 7/2/47, treatment initiated on 7/21 or 7/22/47 and terminated on 7/31/47. The animals were killed on 8/1/47. Cultures and agglutination tests were made on blood drawn at autopsy, qualitative cultures from liver, spleen and kidneys, and quantitative cultures from weighed portions of the spleen. Among the animals infected with 12 organisms, only 2 of the 6 controls and 1 of the 6 treated animals were found to be infected at autopsy. Among the animals infected with 12,000 organisms, 4 of the 5 controls which survived the test period and all of the 6 treated animals were found to be infected. No quantitative differences were apparent in the number of organisms present in the spleens of treated or untreated animals.

Experiment 2. A second lot of polymyxin assaying 900-950 units per mg. was employed in a second experiment. Twenty-five guinea pigs weighing about 400 grams each were divided into 3 groups. Group A, consisting of 8 animals, was treated with 1.5 mg. of polymyxin per animal every 4 hours for 10 days, group B made up of 6 animals was given 1.25 mg. of polymyxin and 2.5 mg. of streptomycin every 4 hours for ten days, and group C consisting of 11 animals, served as controls. The animals were inoculated intraperitoneally with 3,400 organisms of *Br. suis* strain 6 on 9/6/47. The animals were weighed and treatment started on 10/1/47, continuing until 10/11/47. The drugs were given intramuscularly for the first six days and subcutaneously for the remaining 4 days. On 10/17/47, the animals were bled and killed with chloroform. The weight at time of autopsy and the gross pathological lesions were recorded. Agglutination titers against *Br. abortus* were determined for each serum, qualitative cultures were made from liver, spleen, kidney, and blood, and quantitative cultures from weighed portions of spleen. Eight animals failed to survive the test period and are excluded from the final results, which were based on 4 animals surviving in group A, 5 in group B, and 8 in group C. All animals were demonstrated to be infected and no differences were apparent in the degree of infection among animals in the three groups.

Experiment 3. Polymyxin obtained from the J. T. Baker Chemical Company and containing 6000 u. per mg. was tested in this experiment.

Forty-two guinea pigs weighing about 350 grams were inoculated subcutaneously with 950 *Br. suis* organisms strain 6 on 4/21/48. The animals were weighed and divided into the following groups on 5/5/48 at the time treatment was started: Group A was given a total of 5 mg. of polymyxin per kg. of body weight per day subcutaneously in 6 divided doses, Group B was given 100 mg. of sulfadiazine per kilogram of body weight once each day orally and 30,000 units of streptomycin per kg. of body weight per day subcutaneously in 6 divided doses, and Group C was given no treatment. Treatment was discontinued on 5/19/48. At autopsy on 6/2/48, agglutination tests were performed on serum from blood drawn at this time, gross lesions observed, the spleens weighed, and qualitative cultures made from liver and spleen. The results obtained are summarized in TABLE 1. Ad-

TABLE 1

THE EFFECT OF POLYMYXIN AND OF COMBINED SULFADIAZINE AND STREPTOMYCIN IN THE TREATMENT OF GUINEA PIGS INFECTED WITH *Brucella suis*

Group	Number of animals	Therapy	Daily dosage* per kg. body wt.	Ave. wt. before therapy	Ave. wt. at autopsy	Ave. wt. loss or gain	Ave. wt. of spleen	No. of positive spleen cultures	No. of positive liver cultures	Mean agglutination titer	Per cent infected
				gm.	gm.	gm.	gm.				
A	11	Polymyxin	5	367	353	-14	3.21	10†	9	1:1280	100
B	12	Streptomycin and sulfadiazine	30,000	324	401	+77	0.76	2	0	1:40	16.6
			100								
C	14	None	—	336	349	+13	2.91	13	10‡	1:640	92.8

* Dosage of polymyxin in mg., of streptomycin in units, of sulfadiazine in mg.

† One spleen culture contaminated but liver culture positive.

‡ One animal with negative cultures from liver and spleen and negative agglutination titer.

ministration of polymyxin in the dosage stated above had no effect upon the course of experimental *Br. suis* infection in guinea pigs. Among the animals treated with polymyxin, the cultures from the spleen of one animal were contaminated, but cultures of liver yielded *Br. suis* and the agglutination titer of serum obtained at autopsy was 1:320 against *Br. abortus*. One animal in the control group was entirely normal. Only 2 isolations were made from the spleens of animals treated with sulfadiazine and streptomycin.

Experiment 4. *P. tularensis* strain R was cultured on glucose cystine blood agar and incubated at 37°C. for 24 hours. The resultant growth was removed, suspended in 0.85 per cent salt solution, and washed by repeated centrifugation and resuspension in salt solution. Serial ten-fold dilutions were made of the final suspension of organisms and the infective titer determined by injection of 0.3-cc. amounts of dilutions from 10^{-6} to 10^{-10} intraperitoneally into groups of 5 mice each. The mice employed weighed

15 grams. All animals given 0.3 cc. of the 10^{-8} dilution of the suspension of organisms died, 3 of 5 of those given the 10^{-9} dilution succumbed, and none of those given the 10^{-10} dilution died. A group of 37 mice was inoculated intraperitoneally with 0.3 cc. of the 10^{-7} dilution of the suspension of *P. tularensis*. Seventeen of these mice were treated with polymyxin and 20 served as controls. Treatment with polymyxin (900–950 u. per mg.) was started 24 hours after administration of the infective dose of organisms and consisted of 0.125 mg. administered subcutaneously twice daily. This corresponds to a daily dose of about 166 mg. per kg. of body weight. All treated and control animals were dead 120 hours after being infected. Thus, the drug had no effect upon experimental tularemia.

Discussion

The status of the types of antibiotics isolated from *Bacillus polymyxa* was by no means clear when these studies were initiated. Other papers presented in this monograph show that polymyxin is only one of the agents which has been isolated from *B. polymyxa*. Aerosporin has also been isolated and appears to consist of a number of different materials. Consequently, the data which we have recorded must be interpreted on the basis of the sources from which the compounds were obtained. Since the amounts of polymyxin were limited and since little information was available concerning the pharmacology of the drug, the dosages of polymyxin administered to the animals were arrived at empirically. *In vitro* tests for the sensitivity of *Br. abortus* to streptomycin and polymyxin indicated that streptomycin inhibited the test organism in a final concentration of 0.06 μ g. per ml. and polymyxin at levels of 0.5 to 2.5 μ g. per ml.

Polymyxin in the amounts employed had no recognizable influence upon the course of infection produced by *Br. suis* in guinea pigs. The largest amount of polymyxin, a total of 9 mg. daily (25.7 mg. per kg. of body weight) given subcutaneously in 1.5 mg. amounts every 4 hours, had no effect upon the experimental infection. Combined therapy consisting of 1.25 mg. polymyxin and 2.5 mg. of streptomycin, administered 6 times daily for 10 days, likewise yielded negative results. Since the first two experiments were of such a negative character, it was considered possible that the technique of testing for the antibiotic activity of polymyxin was too severe and an additional group of animals treated with streptomycin and sulfadiazine was included in the last experiment. As only 2 of 12 animals treated by this method were infected at time of autopsy, it seems apparent that the technique employed was not too stringent.

Summary

Three lots of polymyxin tested for therapeutic activity in guinea pigs experimentally infected with *Brucella suis* failed to alter the course of infection.

In a single experiment, polymyxin was tested and shown to be incapable of preventing deaths among mice infected with *Pasteurella tularensis*.

References

1. BENEDICT, R. G. & A. F. LANGLYKKE. 1947. Antibiotic activity of *Bacillus Polymyxa*. J. Bact. **54**: 24-25.
2. STANSLY, P. G., R. G. SHEPHERD, & H. J. WHITE. 1947. Polymyxin: a new chemotherapeutic agent. Bull. Johns Hopkins Hosp. **81**: 43-54.
3. AINSWORTH, G. C., ANNIE M. BROWN, & G. BROWNLEE. 1947. Aerosporin; an antibiotic produced by *Bacillus aerosporus* Greer. Nature **160**: 263.
4. BROWNLEE, GEORGE & S. R. M. BUSHBY. 1948. Chemotherapy and pharmacology of Aerosporin: a selective gram-negative antibiotic. Lancet **1**: 127-132.
5. SWIFT, P. N. 1948. Treatment of pertussis with Aerosporin. Lancet **1**: 133-135.

THE CLINICAL USE OF POLYMYXIN*

By EMANUEL B. SCHOENBACH, MORTON S. BRYER, AND PERRIN H. LONG

Department of Preventive Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland

Polymyxin is an antibacterial agent uniquely active against several species of gram-negative microorganisms. Adequate protection of experimentally infected animals was demonstrated with doses of the antibiotic far smaller than those at which any toxic manifestations were noted. Trial in patients, therefore, appeared warranted. These early cases cannot be considered therapeutic trials to establish the clinical use of this antibiotic, but rather as early explorations into the pharmacology and human host-drug relationships.

The small series of patients treated thus far, the varied types of infections, and the great age range of these patients render any generalization or statistical treatment untenable. Gram-negative infections differ markedly from many gram-positive infections in that complications often exist which aid in the initiation of the disease or its maintenance. Many of the cases treated were suffering from an underlying dermatitis, neoplasm, burn, etc. Thus, the clinical course of the untreated case is more difficult to predict, and that of the treated case, to evaluate. Resort to bacteriological procedures coupled with objective and subjective clinical observations of each case is necessary. The observations on these patients will be presented, and data as to tolerance, blood levels attained, urinary excretion, untoward reactions, and cultural examinations reported. The *course* of the patient while on polymyxin therapy may be followed, but no definite deductions as to cause and effect will be attempted.

The dosages of polymyxin employed in these trials varied. Some cases were treated with a daily dose of 3 mg./kg. per day, divided into 8 equal amounts and given at 3-hour intervals. Others were given a daily dose of 7 mg./kg./day. In these patients, an initial dose of 3 mg./kg. followed by two doses of 2 mg./kg. at 8-hour intervals was employed, and thereafter the drug was continued at equal amounts given three times a day. This dosage was not tolerated for more than three to four days and then was often reduced to 3 mg./kg. per day. A few cases have been treated with doses of 4 mg./kg. per day, given as four divided doses at 6 hour intervals. All drug was dissolved in a phosphate buffer pH 7.4 and administered *via* the intramuscular route. A concentration of 60 to 120 mg. in two cubic centimeters was tolerated locally. At the higher concentrations, dull drawing pain was noted for several minutes following the injection. In several individuals, after repeated injections, small, reddened, tender areas persisted at the injection site. However, almost all patients preferred this discomfort to two doses of more dilute material. The injections were not discontinued in any patient because of local reactions. On one occasion,

* These investigations were supported by grants received from the Abbot Laboratories, Lederle Laboratories Division of the American Cyanamid Company, Parke, Davis and Company, and the Upjohn Company.

subcutaneous administration was employed without discomfort. No intravenous administration was attempted.

All the solutions were freshly prepared before use. The polymyxin was obtained as a vacuum dried cake in sealed ampules containing 20 mg. of the hydrochloride salt. Three different lots prepared at the Lederle Laboratories Division of the American Cyanamid Company were employed. These preparations are said to contain approximately 70 per cent pure antibiotic.

Patients have been treated with polymyxin for periods up to 15-20 days. Most patients received the drug for 5-10 days, but in a few, because of urinary changes, this form of therapy was discontinued earlier.

The blood levels attained when treatment was given every three hours at a dosage level of 3 mg./kg./day were 0.6 to 1.3 micrograms per cc. of serum within 12 hours after therapy was begun. On continued treatment, these pre- and post-injection levels rose to 2.5 and 5.0 micrograms per cubic centimeter on the 5th and 10th day. When larger doses were administered, the levels attained were higher and a larger variation between pre- and post-injection periods was noted. When seven milligrams per kg. per day were given, divided into three daily doses, levels of 5 to 10 micrograms were noted and these would rise to over 20 micrograms one hour after an injection of 2 mg. per kg. of body weight. Accumulation of drug was not as apparent in this group however. This may be due to the inaccuracy and inherent difficulties in its determination. A biologic titration method was used employing *E. coli* as the test organism. Because of the contamination of the urine with gram-positive species which was often encountered, and of the inhibition of *E. coli* by fresh serum, it was decided to heat all specimens to 56°C. for $\frac{1}{2}$ hour. Filtration had been found to remove the major portion of the polymyxin. Polymyxin was also noted to deteriorate on standing in alkaline solution. A certain degree of loss with our procedure was inevitable even though the control standard solutions were similarly heated, either in broth or serum. The reported concentrations of polymyxin must therefore be considered as only approximate.

No polymyxin was detected in the cerebrospinal fluid after its intramuscular administration in one patient who was treated for four days. This patient had a purulent meningitis, in which the blood-brain barrier is usually more permeable than normal. In animals, no polymyxin has been detected in the cerebrospinal fluid following intramuscular administration. When 2.0 milligrams of polymyxin was introduced intrathecally at the level of L₄, 5.0 micrograms per cc. was detected in the spinal fluid obtained 12 hours later by cisternal puncture.

When polymyxin in daily dosages of 4-7 mg./kg. was administered to patients, a lag in its urinary excretion of approximately 12 hours was noted. Drug then appeared in the urine, increasing rapidly, so that by 72 96 hours approximately 60 per cent of the administered drug had been excreted. The concentration of polymyxin in the urine varied between 10 and 160 micrograms per cubic milliliter. No clearance studies have been performed.

Untoward reactions to polymyxin have been observed. The outstanding

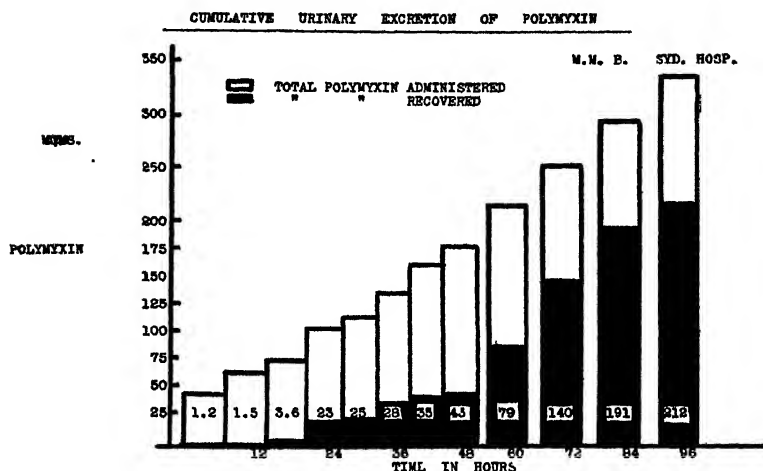


CHART I.

complication has been albuminuria, cellular casts, granular casts, and large epithelial cells which appeared after 72-96 hours of treatment. This has been noted in the majority of individuals treated, especially with the larger infrequently administered doses. The albuminuria was heralded by a change in pH of the urine from acid to alkaline reaction. The specific gravity of the urine approached 1.010. In many of the patients, the albuminuria disappeared as treatment was continued. Azotemia was noted in four patients and marked oliguria was observed in one.

Another minor reaction observed was possible drug fever after 10 days of therapy in an individual with chronic brucellosis. Epigastric distress, with anorexia, unassociated with nausea or vomiting, which persisted for a few hours a day and which disappeared while treatment was being continued, was noted on five occasions in four patients. A peculiar histamine-like flush following an injection was seen twice in the same patient. One further observation, that the temperature sometimes seemed to persist above 99°F. while the drug was being given, only to fall to normal upon its discontinuance, may be added. Repeated blood counts, liver function tests, and careful questioning have not elicited other reactions of note. No skin rashes were observed. One patient developed an eosinophilia of 8-12 per cent without any other signs of allergic reaction.

Four patients treated with polymyxin have died. One was a case of typhoid fever treated in the fourth week of illness for 2 days with a daily dose of 4 mg./kg. He was in circulatory shock, distended, and delirious when treatment was initiated. Sixty hours after polymyxin treatment was instituted, he died with jaundice, albuminuria, and hypopotassiumemia. The second case was a patient with *Aerobacter aerogenes* infection, carcinoma of the prostate, and lymphosarcoma of the bladder. He responded well to polymyxin therapy, blood culture and temperature becoming normal. His blood urea nitrogen on admission was 95 mg. per 100 ml.,

and he finally died in uremia. *Post mortem* examination revealed bilateral obstruction of the ureters by neoplasm and diffuse pyelonephritis. The third case was a 58-year-old male who had a septicemia, right upper lobe pneumonia, osteomyelitis, meningitis, and a large retroperitoneal abscess. He had been ill for four months and had been treated with sulfonamide drugs, several courses of streptomycin, and penicillin. The infecting organism, *K. pneumoniae* Type B, was resistant to 320 micrograms of streptomycin per milliliter *in vitro*. Its *in vitro* sensitivity when tested with polymyxin was 0.3 micrograms per milliliter. His fever fell to 100°F., blood cultures became negative, and the spinal fluid showed some improvement while on intramuscular polymyxin therapy. On the fourth day, however, following a lumbar puncture, he went into shock. The spinal fluid obtained was thick green pus. The patient died soon thereafter. *Post mortem* examination revealed a huge retroperitoneal abscess which had eroded the intervertebral disc between L₅ and S₁ and had perforated into the subarachnoid space. There was no evidence of polymyxin toxicity in the pathological sections.

A patient with bacterial endocarditis due to *Aerobacter aerogenes* secondary to the removal of an infected kidney died in congestive heart failure 72 hours after polymyxin treatment was begun.

Clinical course of the patients treated with polymyxin may be summarized according to the type of infection present. Three cases of infection with *Ps. aeruginosa* were treated with polymyxin. Case 1 was a nine-year-old boy with a generalized exfoliative dermatitis which became infected with *Ps. aeruginosa* and *Streptococcus hemolyticus* B. He also had a pyelonephritis with *Ps. aeruginosa* as the infecting organism. He was gravely ill and despite large doses of penicillin, streptomycin, and sulfadiazine did not improve. Local compresses of streptomycin were also ineffectual in ameliorating this infection. The gram-negative bacillus was resistant to penicillin and 50 micrograms of streptomycin per cubic milliliter *in vitro*. Polymyxin therapy was begun on the 20th hospital day after all other medication except penicillin had been discontinued. He was given an initial dose of 20 mg. intramuscularly and 10 mg. every three hours thereafter (3 mg./kg./day). Blood levels of 0.6 to 2.5 micrograms per milliliter of serum were noted. Treatment was continued for 20 days. Cultures of the urine became negative for *Ps. aeruginosa* within 24 hours and skin lesions showed a marked decrease within 72 hours, so that only occasional areas yielded a positive culture. These latter areas were negative for *Ps. aeruginosa* ten days thereafter, and only hemolytic streptococci persisted. Urine examination revealed a 4+ albuminuria and casts on the third day which disappeared while treatment was continued. Blood non-protein nitrogen was 37 mg. per cent before therapy with polymyxin and after twenty days of treatment was 36 mg. per cent. Penicillin therapy was continued because hemolytic streptococci were repeatedly cultured from the skin areas. These lesions gradually cleared.

Case 2 was a 13-months-old infant with a severe third degree burn of the

back, involving 25 per cent of body area, and second degree burns of face, buttocks, thighs, calves, and elbows. A first degree burn of forehead and ears was present. The burned areas were infected with *Ps. aeruginosa*. The child was gravely ill with a temperature of 104°F. and convulsions. Polymyxin was administered intramuscularly every three hours in a daily dose of 3 mg./kg. The child became afebrile, the exudate cleared rapidly, and cultures became negative. Skin grafting could be performed six days after polymyxin was begun. The child had an uneventful convalescence. Polymyxin was given for twelve days without complications in this case.

Case 3 was an 8-weeks-old child admitted to Sydenham Hospital with a diagnosis of diphtheria and laryngeal obstruction. The diagnosis of diphtheria was not confirmed by cultures, but a tracheotomy was required. Several months later, the tracheotomy wound became infected with *Ps. aeruginosa*. Polymyxin was administered intramuscularly 7½ mg./kg./day divided into 3 doses. Albuminuria and casts were noted on the third day of therapy and the drug was discontinued. The wound infection cleared rapidly.

Five cases of pertussis were treated with polymyxin for 3-5 days. The dosage employed varied between 3-7½ mg. per kg. per day. The results are summarized in TABLE 1. The children under one year appeared to benefit from therapy. The older age group was more difficult to appraise.

Three cases of *Acrobacter aerogenes* bacteremia were treated. Blood cultures became negative in two within twenty-four hours after onset of therapy and the patients soon became afebrile. One patient with endocarditis had a marked reduction in the number of colonies and reduction in fever, but died in congestive heart failure soon after therapy was begun. One case of peritonitis secondary to a perforated appendix was treated with polymyxin. The temperature fell to normal after three days and the patient made an uneventful convalescence.

Two cases of acute brucellosis were treated with polymyxin for 10-14 days. Both experienced relief of symptoms rapidly and became afebrile within 3 to 4 days. On follow-up several months later, both were asymptomatic. No untoward reaction was noted except for a transitory albuminuria in one of the patients.

A 45-year-old male with a low-grade fever and repeatedly positive cultures for *Brucella suis*, despite sulfadiazine and streptomycin therapy, was treated with polymyxin. Blood cultures became negative within 36 hours after therapy was begun. Treatment was stopped after 10 days. Blood cultures remained negative although a low grade fever to 100°F. persisted. The patient felt improved symptomatically. Twenty-six days after polymyxin therapy was discontinued, he developed chills and high fever. Two blood cultures were positive for *Brucella suis*. He is now under treatment again. This patient developed an albuminuria, casts, and fixation of urinary specific gravity at 1.008-1.012 four days after polymyxin was begun. The albuminuria and casts disappeared despite the continuation of polymyxin. However, the blood urea nitrogen rose from 19 to 35 mg. per cent. The azotemia disappeared when polymyxin was discon-

TABLE I
 SUMMARY OF PERTUSSIS CASES

Pt.	Age	Day of disease	Polymyxin			Results ⁴				Remarks	
			Dose daily	Days	Blood level	Before R _x		After R _x			Clinical
						np	cp	np	cp		
			mg./kg.		γ/cc.						
G. B.	6 wk.	4	3	5	2.0	—	—	—	—	Temp. normal, cough disappeared 36 hours	Extremely ill child
J. B.	14 mo.	7†	3	5	4.0	—	—	—	—	Temp. normal 24 hrs. Cough diminished, no whoop after 2 days	Not severely ill before treatment. Whooping for 7 days
C. M.	7 mo.	10	7½	5	2.0	+	+	+	+	Decrease in cough. No parox. after 3 days	Severe whoop with convulsions. Albuminuria and casts 3rd day
J. H.	4 yr.	11	7½	5	2.0	+	+	+	—	Decrease in symptoms after 7th day	Had bronchopneumonia and developed tuberculosis. Effect of drug questionable. Developed albuminuria 4th day
M. J.	2 mo.	7	7½	3		—	—	—	—	Cough less and no vomiting after 2 days	Severely ill. Developed albuminuria and casts on 3rd day

* Np = nasopharyngeal culture; cp = cough plate.

† After onset of cough.

tinued, and subsequent renal studies have been normal as measured by PSP excretion, urinary concentration and dilution tests, blood urea determination, and microscopic examination of the urine.

A 47-year-old farmer with chronic brucellosis for several years has been given two courses of polymyxin. Symptomatically, no improvement has

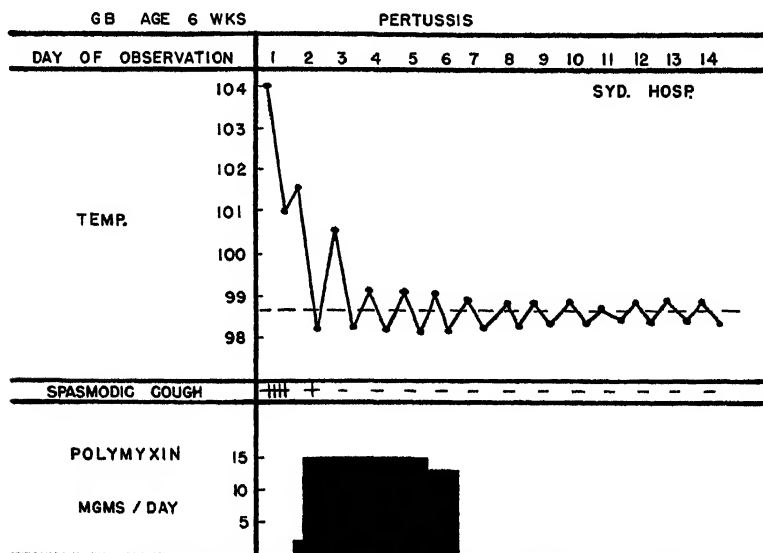


CHART II.

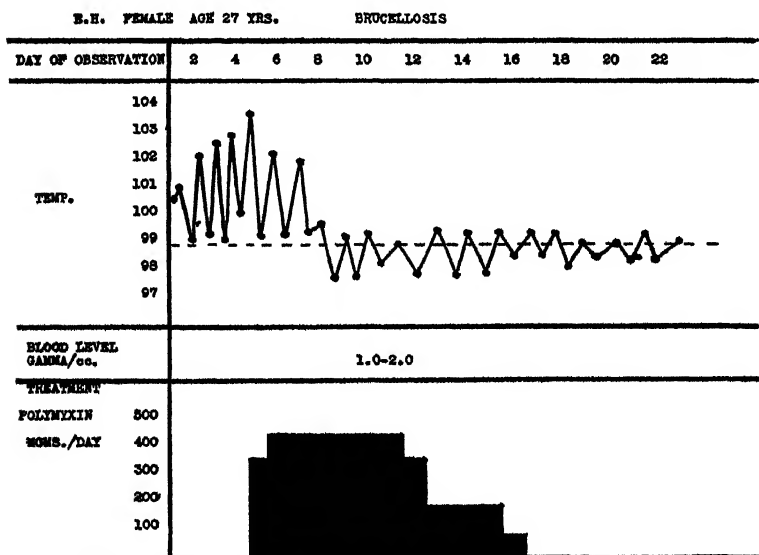


CHART III.

been noted. Blood cultures were not positive either before or after treatment. During the first course of treatment, his temperature subsided, only

to recur on the 10th day. A drug reaction was suspected and the fever disappeared when the drug was discontinued. Subsequent trial of the drug caused a return of fever which subsided when it was again discontinued. The patient has continued to have recurring symptoms and febrile episodes, and it is thought that polymyxin was without value in this patient.

Two cases of infection with *K. pneumoniae* have been treated. The first case, who died, has been presented earlier. The second case had a chronic bronchiolitis and pneumonia which had been present with remissions and exacerbations for three years. A gram-negative bacillus classified as *K. pneumoniae* had been persistently recovered from the sputum. Treatment with sulfonamides, penicillin, and streptomycin had repeatedly been attempted. The initial trial of streptomycin was associated with marked clinical improvement one year ago. However, the drug had to be discontinued and subsequent administration was unassociated with any

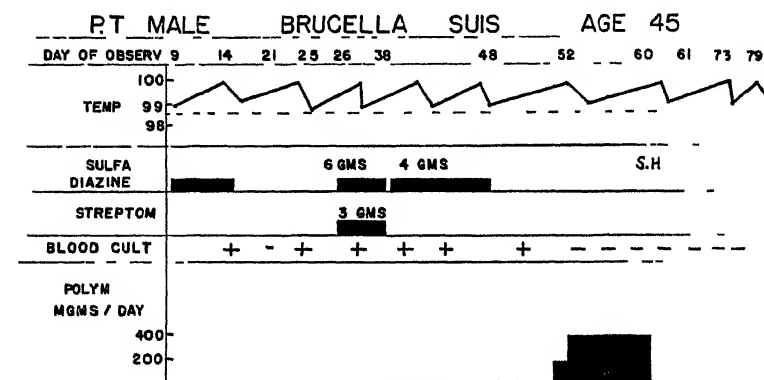


CHART IV.

clinical response. Polymyxin in doses of 3 mg./kg./per day was administered for 15 days intramuscularly. Aerosol inhalations four times a day, of 5 cc. of polymyxin solution (0.2 milligrams per cc. of distilled water), were given during the same period. The sputum rapidly diminished from 4 ounces per day to less than $\frac{1}{2}$ ounce per day. Cultures of the sputum showed fewer organisms with a disappearance of gram-negative bacilli in smears and cultures after 10 days of treatment. The temperature became normal, and serial x-rays revealed clearing of the infiltrations in the left upper lobe of the lung. This patient developed a mild albuminuria with cellular casts while on polymyxin therapy, but no azotemia. The urinary changes cleared when polymyxin was discontinued. Of interest was the development of 8-12 per cent eosinophilia in the blood while on polymyxin therapy. This had not been noted during the previous three years of careful observation. No other allergic manifestations were noted. The organism was sensitive to 0.01 microgram of polymyxin per cubic centimeter. Numerous blood serum levels while on therapy were estimated. They ranged between 1.0 and 1.5 micrograms per milliliter of serum.

Three cases of typhoid fever and one persistent biliary carrier have been given polymyxin. An eleven-year-old male, ill for two weeks, was admitted to the Harriet Lane Home of the Johns Hopkins Hospital. Daily temperatures ranged between 104° and 105°F. Five blood cultures in three days were all positive for *E. typhosa*. Widal was negative. Stool cultures were all positive for *E. typhosa*. Polymyxin was given intramuscularly only. The initial dose was 120 milligrams (3 mg./kg.) followed by 80 mg. three times a day. Because of lack of drug, this was decreased to 60 mg. on the third day of therapy and discontinued on the fifth day. Blood levels of 5.0 to 10.0 micrograms per cubic milliliter of serum were noted on this dosage schedule.

Blood cultures 36 hours after treatment was begun were sterile and remained sterile for more than three weeks. Stool cultures became negative for *E. typhosa* within one week and remained repeatedly negative for three weeks. The temperature fell to normal on the fourth day and after a few elevations to 102°F. again returned to normal limits. Three weeks after polymyxin had been discontinued, fever returned, blood cultures and stool cultures became positive for *E. typhosa*, and the patient experienced a relapse of disease.

This patient had a 2+ albuminuria and fixation of urinary specific gravity at 1.010 upon admission. While on polymyxin therapy, a 3+ albuminuria was noted with the appearance of occasional cellular casts. An oliguria of approximately 125 cubic milliliters per day was noted. This gradually disappeared during the week after polymyxin was discontinued. The non-protein nitrogen of the blood rose progressively from 19 mg. per 100 ml. of blood to 68-110 mg. per 100 ml. Two weeks after polymyxin therapy had been begun, the non-protein nitrogen was 172 mg. per cent. Blood chlorides were reduced to as low as 89 milliequivalents. The azotemia and hypochloremia gradually disappeared during the next two weeks. The albuminuria has persisted. This case is an example of temporary bacteriological and clinical response complicated by oliguria in a previously damaged kidney. The brother of this patient died and has been discussed earlier in this presentation.

A 62-year-old male admitted to the Johns Hopkins Hospital with right lower lobe pneumonia, phlebitis, prostatic obstruction, and high fever, was treated with penicillin and streptomycin without improvement. He developed a rash on the abdomen, and blood and stool cultures were subsequently reported positive for *E. typhosa*. The patient was gravely ill. Polymyxin in doses of 4 mg./kg. per day was given intramuscularly for five days. Temperature fell to 102°F. within 24 hours and thereafter to normal on the fifth day after treatment was instituted. Blood and stool cultures were negative for *E. typhosa* subsequent to polymyxin therapy, except for one stool culture on the second day of treatment. A transient increase of albuminuria was noted on the fourth day unaccompanied by azotemia. The patient had required an indwelling catheter, and one month later a suprapubic prostatectomy was performed. The patient made an uneventful recovery.

A 27-year-old female had experienced a mild illness during an epidemic of typhoid one year earlier. Her daughter had had clinical typhoid fever. In March, 1948, a cholecystectomy had been performed which resulted in a biliary fistula. Drainage from the fistula was found repeatedly to contain *E. typhosa*. The patient was treated with polymyxin intramuscularly, 120

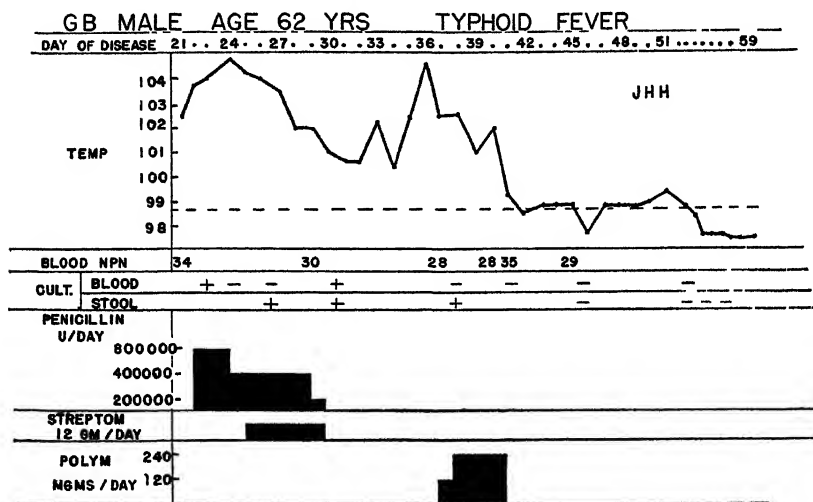


CHART V.

TABLE 2
SUMMARY OF CLINICAL TRIALS WITH POLYMYXIN

Disease	No. treated	Age of patients	Treatment		Results	
			Daily dose mg./kg.	No. of days	Cultural	Clinical
<i>Pseudomonas aeruginosa</i> infection	3	8 wk.-9 yr.	3-7	3-20	Excellent	Excellent
Pertussis	5	6 wk.-4 yr.	3-7	4-5	Doubtful	Good
Peritonitis	1	Adult	3	3	--	Good
<i>Aerobacter aerogenes</i>	3	Adults	4-7	4	Excellent	Good
<i>K. pneumoniae</i>	2	Adults	3-4	4-15	Excellent	Good
Brucellosis	4	Adults	3-7	10-20	Good	Good, acute; chronic, none
Typhoid	4	11, 16, 27, 62 years	4	2-5	Good	Variable

mg. as an initial dose and 60 mg. three times a day thereafter. On the fifth day of therapy, albuminuria, nausea, and epigastric distress was noted. Polymyxin therapy was discontinued with an immediate subsidence of nausea and albuminuria. The biliary cultures became negative 48 hours after therapy was instituted and have remained negative to date.

Summary and Conclusions

A total of twenty-two patients suffering from numerous types of infection caused by the gram-negative bacilli have been treated with polymyxin. The results are summarized in TABLE 2. The chief complication encountered in the course of therapy has been renal toxicity, manifested by albuminuria, cellular casts, a fixation of specific gravity of the urine, and azotemia. In the majority of the cases, the albuminuria has been transitory and the more severe evidence of damage has been limited to those patients who had preexisting renal impairment. Nausea and anorexia with occasional epigastric distress has been observed in four patients. These symptoms have disappeared when treatment was discontinued and, on two occasions, even while therapy was continued. An eosinophilia of 8-12 per cent was noted in the blood of one patient. Probable drug fever was observed in a patient treated for chronic undulant fever. The results with local therapy with polymyxin have not been included in this report.

Polymyxin appears to be a valuable antibiotic for the treatment of infections due to the gram-negative group of microorganisms, and it is hoped that further purification will eliminate the risk of renal damage so that clinical appraisal and application may be extended.

REMARKS ON CLINICAL RESULTS WITH POLYMYXIN A AND B

By GEORGE BROWNLEE

The Wellcome Physiological Research Laboratories, Beckenham, Kent, England

The first clinical report on the use of the selectively Gram-negative antibiotic polymyxin A ("Aerosporin") was the preliminary observations of Swift¹ on the treatment of pertussis. This clinical demonstration was restricted to 10 unselected patients, mostly late in the disease, and complicated by secondary invaders, the organism being isolated in only one instance. Swift's scheme of dosage was between 0.4 to 0.8 mg. per kg., 4-hourly for five days, and was related to a blood level of 0.2 μ g. to 0.4 μ g. per ml.

He was able to conclude that the disease was aborted dramatically in early pertussis, and significantly modified in late disease.

Additional Cases

Pertussis: P. N. Swift, M.R.C.S., M.R.C.P., Paediatric Unit, The County Hospital, Farnborough, Kent.

In a further series of 17 children with pertussis who were treated with 0.5 mg. per kg., Swift has verified his earlier observations. In infants treated early in the disease, the response is dramatic; the older child appears to have the course of the disease shortened. Of the 17 children, 8 were treated with polymyxin B, in which instances proteinuria was not observed.

Pertussis: L. J. M. Laurent, M.D., M.R.C.S., Park Hospital, Hither Green, London.

Laurent treated 17 children with pertussis, from which the causal organism was isolated before therapy, with a dose of 0.5 mg. per kg., 4-hourly for 5 days. Eight of the patients were treated with polymyxin A together with methionine. In spite of this attempt to protect against the nephrotoxic action of polymyxin A, a transient proteinuria developed on the 5th day. Seven patients treated with polymyxin B, showed only the mildest proteinuria of temporary duration. It is Laurent's opinion that the only patient to respond dramatically is the small infant, severely ill with vomiting and paroxysms. This disease aborts in 3 days or less. Half of the total observations were in this category, and the other half showed shortening of the usual course of the disease.

Pertussis: N. M. Coutts, M.B., M.R.C.S., Brook Hospital, Shooter's Hill, London.

In this series of 10 patients with proved pertussis treated with polymyxin B, the dose was 0.5 mg. per kg. 4-hourly for five days. With the exception of a boy of 5 years, who was removed from hospital on the sixth day, all were children of less than one year old, who improved. However, three patients were not treated until the fourth week of disease. In the remaining six children the disease was aborted.

In all but two instances, there was pyrexia during treatment but no

marked constitutional disturbance. There were a few mild local reactions, and one case of albuminuria.

Genito-urinary infections: C. C. Cookson, M.B., F.R.C.S., L.R.C.P., The County Hospital, Farnborough, Kent.

Eight examples of urinary infection with complications were treated with polymyxin B.

Two patients, one a cystitis and epididymitis of *E. coli* and *Ps. aeruginosa* origin, and the second a cystitis and orchitis of *E. coli* origin, were bacteriologically cured after a 3-day scheme of dosage. The infection of two others, one a cystitis with diverticulum of the bladder with *E. coli*, and the other a prostatitis of *E. coli* and *Ps. aeruginosa* origin, were temporarily controlled, with a return of the infection when treatment ceased.

The remaining four patients were of mixed infections with sensitive organisms, *E. coli* and *Ps. aeruginosa*, together with insensitive organisms, *Pr. vulgaris* and a streptococcus. With the exception of one patient with pyonephrosis, polymyxin B eliminated the sensitive organisms.

The dose of polymyxin B was standardized at 20 mg. 4-hourly, for the first four patients for three days, and the remainder for five. The ideal scheme of dosage cannot be determined from this small series. The relapse of the patient with prostatitis due to *Ps. aeruginosa* after 3 days of treatment indicated the need for a more prolonged course.

It may be concluded that parenteral polymyxin B can eliminate sensitive organisms from the genito-urinary tract. Cures may be effected when the causal organism is sensitive and in the absence of anatomical defects. Even in the latter patients, the organisms disappear sufficiently long for surgical interference to be made.

Typhoid Carriers: W. W. Kay, M.D., Ch.B., and C. K. McDonald, M.B., Ch.B., West Park Mental Hospital, Epsom, Surrey.

Preliminary observations which have been made on a small series of four adult typhoid carriers have been useful in establishing the fact that 100 mg. (but not 50 mg.) of polymyxin B given 4-hourly will rapidly eliminate sensitive organisms from the gut. The typhoid carriers have, up to date, all again become excretors after variable periods of time. The oral therapy is now being combined with parenteral.

Gastro-enteritis --Infants: I. A. B. Cathie, M.D., The Hospital for Sick Children, Great Ormond Street, London.

Cathie has observed the effect of an oral dose of 50 mg. daily of polymyxin B given 6-hourly for 4 days, on an unselected series of 20 infants suffering from gastro-enteritis. These trials were arranged in a Latin-square experiment in which the nature of the medicament within the treated groups was unknown to the physician. Of the 20 cases treated with polymyxin B the clinicians picked out 7 which recovered atypically quickly and in some cases dramatically. An additional 2 cases got better quicker than expected but were not so striking. In the remainder, there was no atypical clinical response which the clinician could detect.

Coliform organisms disappeared completely in 5 cases and were markedly

reduced in 8 others. No correlation was observed between those sterilized and those which improved clinically. Two cases relapsed immediately, but one was a case of eczema in which it was expected.

Gastro-enteritis—Infants: P. N. Swift, M.R.C.S., M.R.C.P., Paediatric Unit, The County Hospital, Farnborough, Kent.

Swift has used an oral dose of 4 mg. per kg. every four hours for five days, since this dose usually sterilizes the gut of the Gram-negative flora. His experience has been limited to a series of 18 infants in which drug therapy has been used as an adjunct to a scheme of management. In only one infant, severely ill, was there dramatic recovery. All cases showed clinical improvement. Thus, from the analysis of weight records kept for 3 weeks after treatment, the treated cases are identified by their deviation above the mean and the untreated cases by their deviation below, or deviation by death.

In a further limited series of cases, there appear to be additional advantages in progressively diminishing treatment over a total period of eight days. It appears that the re-establishment of a gram-negative flora represents a critical time in the life of the gastro-enteritis patient.

Gastro-enteritis—Infants: Tom Anderson, M.D., F.R.C.P., Knightswood Hospital, Anniesland, Glasgow.

This observation is restricted to four infants who made an uneventful recovery on a five-day therapy of 2 mg. 4-hourly by mouth. One severely ill infant received in addition 0.2 mg. 4-hourly systemically.

Gastro-enteritis—Infants: Essentially similar effects have been observed at two additional centers with the same doses orally. F. Kane, M.D., Purdysburn Hospital, Belfast, Ireland, has observed a small series of four infants who responded favorably, and J. Smith, D.Sc., M.D., City Hospital, Aberdeen, has commented on the favorable effects in a series of 12 infants.

Gastro-enteritis of various origin—Adults: The experimental observation that 100 mg. (1 mega unit) of polymyxin B given 4-hourly for three to five days will substantially eliminate Gram-negative organisms from the gut, has been pressed into service to treat gastro-enteritis of various origin in the adult.

A patient was successfully treated as judged by clinical recovery and bacteriological elimination of the causal organism, *Sal. typhimurium*.

References

1. SWIFT, P. N. 1948 Treatment of pertussis with Aerosporin. *Lancet* 254: 133.

